NASA SP-431
C.1

LOAN COPY: R
AFWL TECHNI
KIRTLAND AFT

TECH LIBRARY KAFB, NM
0063533

Digital Processing of Remotely Sensed Images

21 MAY 1981
Technical Library
AFWL Sapona
Laboratories

NASA

NASA SP-431
TECH LIBRARY KAFB, NM
0063533

Digital Processing of Remotely Sensed Images

Johannes G. Moik
Goddard Space Flight Center

NASA Scientific and Technical Information Branch 1980
National Aeronautics and Space Administration
Washington, DC

Library of Congress Cataloging in Publication Data
Moik, Johannes G
Digital processing of remotely sensed images.
(NASA SP ; 431)
Includes index.
1. Image processing. 2. Remote sensing.
I. Title. II. Series: United States. National Aeronautics and Space Administration. NASA SP ; 431.
TA1632.M64 621.36'7 79–16727

For sale by the Superintendent of Documents, U.S. Government Printing Office
Washington, D.C. 20402

Preface

Digital image processing has become an important tool of research and applications in many scientific disciplines. Remote sensing with spaceborne and airborne instruments provides images for the study of earth resources and atmospheric phenomena. Medical research and clinical applications use images obtained from X-ray, infrared, and ultrasonic sensors. Electron microscopy yields data concerning the molecular structure of materials. Astronomy uses images taken in the ultraviolet, visible, and infrared radiation ranges. Military reconnaissance relies on the analysis of images.

The common element is that multidimensional distributions of physical variables are represented as images from which useful information has to be extracted. The image processing scientist must be versed in the electro-optics of sensing, in transmission and display technology, in system and probability theories, in numerical analysis, in statistics, in pattern recognition, and in the psychophysics of vision. The new discipline of image science has developed along these lines. The designer of image processing systems is required to know computer systems, man–machine communication, computer graphics, data management, and data-base management systems.

A large number of papers and several textbooks published in recent years demonstrate the rapid growth of image science. The excellent book by Rosenfeld and Kak [1] [1] covers all aspects of processing monochrome images. Andrews [2], Gonzalez and Wintz [3], and Pratt [4] also emphasize digital processing of monochrome images. Andrews and Hunt [5] treat radiometric image restoration, and Duda and Hart [6] deal with scene analysis in the image processing part of their book. Huang and others [7] and Hunt [8] are two of many excellent survey papers. Billingsley [9] and O'Handley and Green [10] summarize the pioneering contributions made at the Jet Propulsion Laboratory. Rosenfeld's review paper [11] contains a large number of references to the image processing literature.

This book was written to assist researchers in the analysis of remotely sensed images. Remote sensing generally obtains images by various sensors, at different resolutions, in a number of spectral bands, and at differ-

[1] References mentioned in the Preface are listed at the end of chapter 1.

ent times. These images, with often severe geometric distortions, have to be combined and overlayed for analysis. Two separate approaches, one based on signal processing methods and the other on pattern recognition techniques, have been employed for the analysis. This book attempts to combine the two approaches, to give structure to the diversity of published techniques, (e.g., refs. [12] to [14]), and to present a unified framework for the digital analysis of remotely sensed images. The book developed from notes written to assist users of the Smips/VICAR system in their image processing applications. This system is a combination of the Small Interactive Image Processing System (Smips) [15, 16] developed at NASA Goddard Space Flight Center and of the Video Image Communication and Retrieval system (VICAR) [17] developed at the Jet Propulsion Laboratory.

The author expresses his gratitude to P. A. Bracken, J. P. Gary, M. L. Forman, and T. Lynch of NASA Goddard Space Flight Center, and R. White of Computer Sciences Corp. for their critical review of the manuscript. The assistance of W. C. Shoup and R. K. Lum of Computer Sciences Corp. in software development and preparation of many image processing examples is greatly appreciated.

Contents

Preface ... iii
1. Introduction ... 1
2. Image Processing Foundations ... 9
2.1 Representation of Remotely Sensed Images ... 9
2.2 Mathematical Preliminaries ... 11
2.2.1 Delta Function and Convolution ... 11
2.2.2 Statistical Characterization of Images ... 12
2.2.3 Unitary Transforms ... 15
2.2.3.1 Fourier Transform ... 17
2.2.3.2 Hankel Transform ... 21
2.2.3.3 Karhunen–Loéve Transform ... 23
2.2.4 Description of Linear Systems ... 23
2.2.5 Filtering ... 24
2.3 Image Formation and Recording ... 34
2.4 Degradations ... 39
2.4.1 Geometric Distortion ... 40
2.4.2 Radiometric Point Degradation ... 42
2.4.3 Radiometric Spatial Degradation ... 43
2.4.4 Spectral and Temporal Differences ... 44
2.5 Digitization ... 44
2.5.1 Sampling ... 46
2.5.2 Quantization ... 49
2.6 Operations on Digital Images ... 52
2.6.1 Discrete Image Transforms ... 52
2.6.1.1 Discrete Fourier Transform ... 55
2.6.1.2 Discrete Cosine Transform ... 59
2.6.1.3 Hadamard Transform ... 60
2.6.1.4 Discrete Karhunen–Loéve Transform ... 60
2.6.2 Discrete Convolution ... 63
2.6.3 Discrete Crosscorrelation ... 65
2.7 Reconstruction and Display ... 67
2.8 Visual Perception ... 69
2.8.1 Contrast and Contour ... 70
2.8.2 Color ... 72
2.8.3 Texture ... 73

3. Image Restoration ... 77
3.1 Introduction ... 77
3.2 Preprocessing ... 77
3.2.1 Illumination Correction ... 78
3.2.2 Atmospheric Correction ... 78
3.2.3 Noise Removal ... 79
3.3 Geometric Transformations ... 89
3.3.1 Coordinate Transformations ... 103
3.3.2 Resampling ... 109
3.4 Radiometric Restoration ... 114
3.4.1 Determination of Imaging System Characteristics ... 119
3.4.2 Inverse Filter ... 120
3.4.3 Optimal Filter ... 121
3.4.4 Other Radiometric Restoration Techniques ... 122
4. Image Enhancement ... 127
4.1 Introduction ... 127
4.2 Contrast Enhancement ... 128
4.3 Edge Enhancement ... 130
4.4 Color Enhancement ... 141
4.4.1 Pseudocolor ... 148
4.4.2 False Color ... 149
4.5 Multi-Image Enhancement ... 158
4.5.1 Ratioing ... 158
4.5.2 Differencing ... 159
4.5.3 Transformation to Principal Components ... 164
5. Image Registration ... 187
5.1 Introduction ... 187
5.2 Matching by Crosscorrelation ... 190
5.3 Registration Errors ... 191
5.3.1 Geometric Distortions ... 192
5.3.2 Systematic Intensity Errors ... 192
5.3.3 Preprocessing for Image Registration ... 194
5.4 Statistical Correlation ... 194
5.5 Computation of the Correlation Function ... 196
6. Image Overlaying and Mosaicking ... 199
6.1 Introduction ... 199
6.2 Techniques for Generation of Overlays and Mosaics ... 199
6.3 Map Projections ... 200
6.3.1 Classes of Map Projections ... 201
6.3.2 Coordinate Systems ... 203
6.3.3 Perspective Projections ... 203
6.3.4 Mercator Projection ... 205
6.3.5 Lambert Projection ... 207
6.3.6 Universal Transverse Mercator Projection ... 208

6.4 Map Projection of Images ... 209
6.5 Creating Digital Image Overlays ... 211
6.6 Creating Digital Image Mosaics ... 212
7. Image Analysis ... 223
7.1 Introduction ... 223
7.2 Image Segmentation ... 223
7.2.1 Thresholding ... 224
7.2.2 Edge Detection ... 225
7.2.3 Texture Analysis ... 232
7.3 Image Description ... 234
7.4 Image Analysis Applications ... 234
7.4.1 Wind Field Determination ... 234
7.4.2 Land-Use Mapping ... 237
7.4.3 Change Detection ... 243
8. Image Classification ... 249
8.1 Introduction ... 249
8.2 Feature Selection ... 253
8.2.1 Orthogonal Transforms ... 254
8.2.2 Evaluation of Given Features ... 259
8.3 Supervised Classification ... 263
8.3.1 Statistical Classification ... 266
8.3.2 Geometric Classification ... 273
8.4 Unsupervised Classification ... 278
8.4.1 Statistical Unsupervised Classification ... 278
8.4.2 Clustering ... 278
8.5 Classifier Evaluation ... 281
8.6 Classification Examples ... 284
9. Image Data Compression ... 293
9.1 Introduction ... 293
9.2 Information Content, Image Redundancy, and Compression Ratio ... 295
9.3 Statistical Image Characteristics ... 296
9.4 Compression Techniques ... 298
9.4.1 Transform Compression ... 298
9.4.2 Predictive Compression ... 303
9.4.3 Hybrid Compression ... 305
9.5 Evaluation of Compression Techniques ... 306
9.5.1 Mean Square Error ... 306
9.5.2 Signal-to-Noise Ratio (SNR) ... 306
9.5.3 Subjective Image Quality ... 307
Symbols ... 309
Glossary of Imaging Processing Terms ... 315
Index ... 321

1. Introduction

Image processing is concerned with the extraction of information from natural images. Extractive processing is based on the proposition that the information of concern to the observer may be characterized in terms of the properties of perceived objects or patterns. Thus, information extraction from images involves the detection and recognition of patterns. Most information extraction tasks require much human interpretation and interaction because of the complexity of the decisions involved and the lack of precise algorithms to direct automatic processing.

The human visual system has an extraordinary pattern recognition capability. In spite of this capability, however, the eye is not always capable of extracting all the information from an image. Radiometric degradations, geometric distortions, and noise introduced during recording, transmission, and display of images may severely limit recognition. The purpose of image processing is to remove these distortions and thus aid man in extracting information from images.

Image processing operations can be implemented by digital, optical, and photographic methods. The accuracy and flexibility of digital computers to carry out linear and nonlinear operations and iterative processes account for the growth of digital image processing. Digital processing requires digitization of the measured analog signals. After processing, the digital data must be reconstructed to continuous images for display.

Image processing must always use some kind of a priori knowledge about the characteristics of objects to be observed and of the imaging system. Otherwise there would be no basis for judging whether a picture is a good representation of an object. Thus, some form of a priori knowledge must be applied to a degraded image to extract information from it. Processing a degraded image may be different depending on whether the source of the image is known. Hence, one kind of a priori knowledge is concerned with intelligence information.

Another kind of a priori knowledge of great importance to image processing is concerned with the physical process of forming an image. This area includes knowledge of object characteristics and of properties of sensor, recording, transmission, digitization, and display systems. For example, the correction of radiometric and geometric distortions that occur in imaging with a vidicon camera requires knowledge of the characteristics of the vidicon tube. All this information is used to reduce the number of variables involved in processing.

The various steps involved in imaging and image processing may be idealized as shown in figure 1.1. This block diagram also introduces the notation used in this publication. The radiation emitted and reflected by an object is represented by a continuous function $f(x, y)$. $f(x, y)$ is attenuated by the intervening atmosphere to the apparent object radiant energy $f^*(x, y)$ in the sensor's field of view. In the image formation process, the apparent object radiant energy is transformed by a linear optical system into image radiant energy $g_0(x, y)$. The image radiant energy is sensed and recorded with noise by a sensor s to form the recorded image g. The recorded image is digitized to the image matrix $(g(i, j))$ by the operator q, which represents sampling and quantization. Digitization introduces a spatial degradation due to sampling and adds quantization noise. After transmission and processing, which use a priori information, the digital images are reconstructed to continuous pictures and maps and are displayed on cathode ray tubes (CRTs) or photographic transparencies and prints or are printed with ink on paper. Reconstruction introduces another spatial degradation.

This book discusses the techniques employed in the analysis of remote sensing images. Remote sensing is the acquisition of physical data of objects or substances without contact. In Earth-related remote sensing, the objects may be part of the Earth's surface or atmosphere. Because the chemical and physical properties of substances vary, they reflect or emit unique spectra of electromagnetic energy, dependent on time and spatial location. Remote sensing derives information by observing and analyzing these variations in the radiation characteristics of substances. Spatial, spectral, and temporal variations of electromagnetic radiation and polarization differences can be observed.

Spatial variations are the source of brightness differences between elements in a scene and permit recognition of objects through contrast, texture, and shape. Spectral variations or variations of radiant energy with wavelength produce color (false color in the nonvisible region of the spectrum). This characteristic permits recognition of objects by measuring the radiant energy in different wavelengths. Temporal variations of radiation from objects caused by seasonal or environmental changes or manmade effects provide additional information for recognition. Polarization differences between the radiance reflected or emitted from objects and their background may also be used for recognition.

Remote sensing imaging instruments onboard aircraft or spacecraft detect and measure electromagnetic energy at different wavelengths and convert the resulting signal into a form perceivable by the human visual system. The imaging instruments range from conventional photographic cameras, over television systems and optical-mechanical scanners, to fully electronic scanning systems with no moving parts. The choice of instruments is influenced by their ability to detect energy at desired wave-

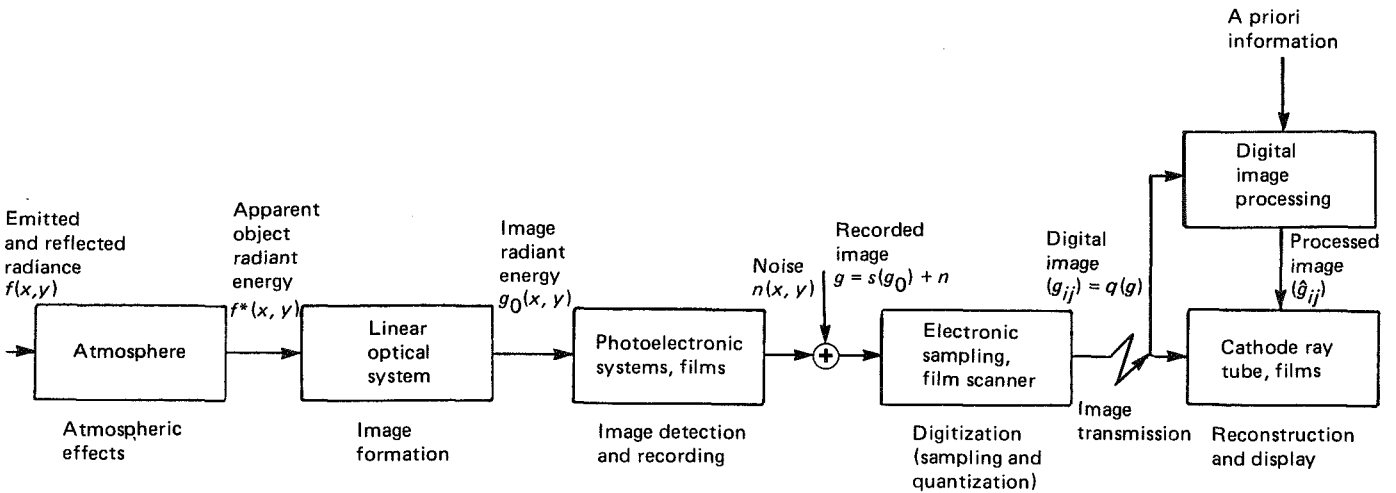


FIGURE 1.1—*Block diagram of imaging process.*

lengths. An introduction to remote sensing instrumentation and an overview of digital image processing activities in remote sensing are given in [18] and [19], respectively.

Because of its global and repeated measurement capability, remote sensing is able to make important contributions to the solution of problems in weather, climate, and Earth resources research. Earth resources applications include detection of changes in previously mapped areas, water resources and land-use monitoring and management, geologic mapping and exploration for geologic resources, detection of crop diseases, and agricultural yield forecasting. Weather applications include the determination of wind fields, temperature and humidity profiles, cloud heights, and severe-storm analysis.

A general block diagram of a remote sensing system is shown in figure 1.2. The basic parts of the system are the scene to be imaged, the airborne or spaceborne imaging and transmission system, the receiving ground station, and the data analysis system. The transmission of remote sensing images may be performed before or after digitization. For example, Landsat Return Beam Vidicon (RBV) images are transmitted in analog form and digitized before processing, whereas Landsat Multispectral Scanner (MSS) images are digitized at the sensor output and are transmitted in digital form.

The basic system parameters are the number of spectral bands, and the spectral, radiometric, and spatial resolution required to discriminate the various objects in a scene. These parameters and the size of the scene to be imaged determine the data rate and processing techniques. The amount of data may be reduced before transmission by coding or preprocessing that selects only useful data for transmission.

In contrast to images taken at the ground, remote sensing images also contain the effects of the atmosphere. The atmosphere absorbs, emits, and scatters radiation. Thus, the information transmitted from the scene

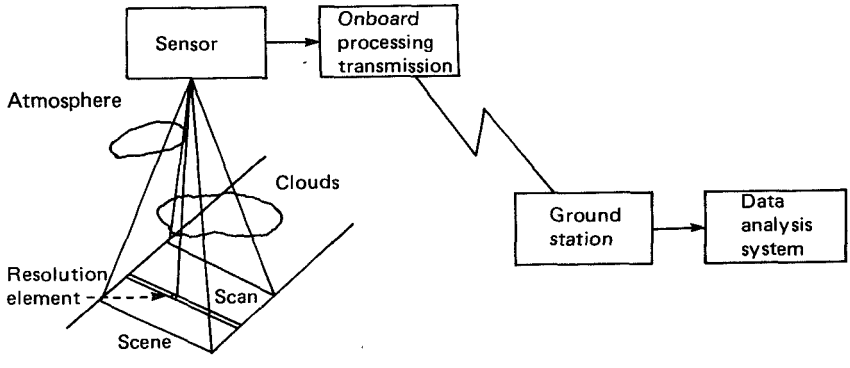


FIGURE 1.2—*Block diagram of remote sensing system.*

to the sensor is attenuated and distorted. The radiant energy scattered and emitted diffusely by the atmosphere into the field of view of the sensor adds noise to the signal. The atmosphere is transparent enough for remote sensing only in small bands of the electromagnetic spectrum, which are called windows. The principal windows lie in the visible, infrared, and microwave regions of the spectrum.

An organization using remote sensing images is not only confronted with image analysis but also with the problem of how to incorporate the acquired and analyzed data into a data-base management system. Without the capabilities of storing and retrieving data by attribute values and relationships between data attributes and integrating the data with ground information, the effective use of remote sensing data will be limited [20, 21]. Successful information extraction from remotely sensed images requires knowledge of the data characteristics, i.e., an understanding of the physics of signal propagation through the atmosphere, of the image formation process, and of the error sources.

Digital image processing techniques can be divided into two basically different groups. The first group includes quantitive restoration of images to correct for degradations and noise, registration for overlaying and mosaicking, and subjective enhancement of image features for human interpretation. The required operators are mappings from images into images. The second group is concerned with the extraction of information from the images. This area of image analysis includes object detection, segmentation of images into characteristically different regions, and determination of structural relationships among the regions. Operators of this group are mappings from images into descriptions of images. These operators convert images into maps, numerical and graphical representations, or linguistic structures.

The major branches of digital image processing are shown in figure 1.3. Because most distortions are nonlinear, digital processing with its pre-

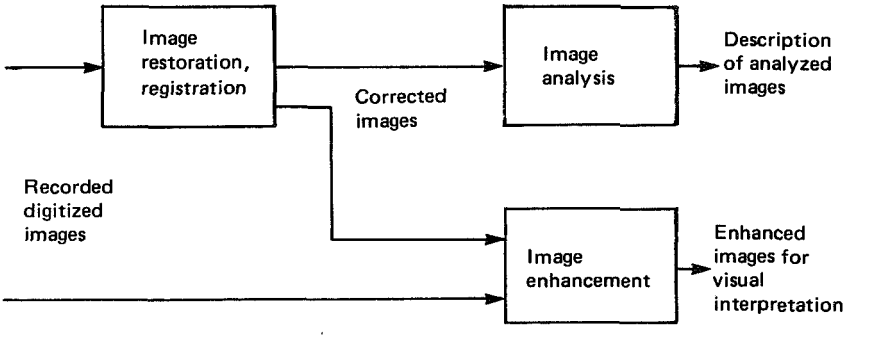


FIGURE 1.3—*Block diagram of digital image processing steps.*

cision and its flexibility to implement nonlinear operations is currently the only feasible technique to solve restoration and registration problems.

Digital image processing requires a system that provides a set of functions for image processing, data management, display, and communication between analyst and system. The functional requirements of such a system can be determined from an understanding of image formation, of recording and display processes, and of image analysis applications and by considering the techniques and strategies applied by a human analyst in the process of information extraction. An effective and convenient language is needed for the analyst to express his processing steps in terms of these functions.

Image analysis with a digital computer in a batch mode requires specification of all processing steps before starting the computer run. Results can only be displayed and evaluated after completion of the run. Therefore, many processing runs may be required to establish the correct analysis procedure. This limitation stretches the period for performing a careful analysis over a long time. The inconveniences and delays often prevent an analyst from exploring all interpretation techniques.

Image analysis is often a heuristic process, and display of the results of intermediate processing steps makes the analysis more comprehensible. Therefore, the analyst should be included in the image processing system. He is the specialist who is capable of integrating the information available in the image with his experience from other data. The combination of man and computer in an interactive system leads to the solution of image analysis problems that neither could solve efficiently alone. The analyst uses an interactive terminal to direct the analysis by means of function keys, alphanumeric keyboard, and light pen or joystick. Results are immediately displayed as images or in graphical form on a display screen and, combined with experience, can be used for selecting the next operation. Thus, the short time between specification of a problem and the return of intermediate results permits a more intelligent choice and sequencing of operations applied to the images. Two interactive image processing systems used for remote sensing data analysis are described in references [16] and [22].

The efficient communication between man and computer requires a language adapted to the problems of the analyst, who is usually not a programmer. Only the operation of the interactive analysis terminal and not of the computer system has to be learned. This simplification reduces errors and permits concentration on the content of the dialogue rather than on its form. Operation of the system in a batch mode should, however, be possible. An interactively determined sequence of processing steps can often be applied to images from particular problems without human interaction.

This book is organized into nine chapters. Chapter 2 introduces digital image processing concepts and reviews mathematical tools from the fields of linear system theory, statistics, theory of random fields and unitary transforms that are the basis for techniques described in later chapters. This review is in no way complete, and the references provide completeness and proofs. Chapter 3 treats image restoration, the correction of geometric and radiometric distortions in images. Chapter 4 deals with image enhancement, the subjective improvement of image quality and appearance. Chapter 5 discusses image registration, the matching of images of the same object scene. Chapter 6 is devoted to overlaying and mosaicking of images. Chapters 7 and 8 discuss image segmentation, the conversion of images into maplike descriptions and image classification, the specific segmentation using pattern recognition techniques. Chapter 9 provides an overview of image data compression techniques. The pictures used as examples are digital images processed with Smips/VICAR [15] at the Goddard Space Flight Center.

REFERENCES [1]

[1] Rosenfeld, A.; and Kak, A. C.: Digital Picture Processing. Academic Press, New York, 1976.
[2] Andrews, H. C.: Computer Techniques in Image Processing. Academic Press, New York, 1970.
[3] Gonzalez, R. C.; and Wintz, P.: Digital Image Processing. Addison-Wesley, Reading, Mass., 1977.
[4] Pratt, W. K.: Digital Image Processing. Wiley-Interscience, New York and Toronto, 1978.
[5] Andrews, H. C.; and Hunt, R.: Digital Image Restoration. Prentice Hall, Englewood Cliffs, N.J., 1977.
[6] Duda, R. O.; and Hart, P. E.: Pattern Classification and Scene Analysis. Wiley-Interscience, New York and London, 1973.
[7] Huang, T. S.; Schreiber, W. F.; and Tretiak, O. J.: Image Processing, Proc. IEEE, vol. 59, 1971, pp. 1586–1609.
[8] Hunt, B. R.: Digital Image Processing, Proc. IEEE, vol. 63, 1975, pp. 693–708.
[9] Billingsley, F. C.: Review of Digital Image Processing, Eurocomp 75, London, Sept. 1975.
[10] O'Handley, D. A.; and Green, W. B.: Recent Developments in Digital Image Processing at the Image Processing Laboratory, Proc. IEEE, vol. 60, 1972, pp. 821–828.
[11] Rosenfeld, A.: Picture Processing, 1973, Comp. Graph. Image Proc., vol. 3, 1974, pp. 178–194.
[12] Proceedings of International Symposia on Remote Sensing of the Environment. University of Michigan, Ann Arbor, Mich.
[13] Proceedings of Symposia on Machine Processing of Remotely Sensed Data. Purdue University, Lafayette, Ind.

[1] These references include references mentioned in the preface.

[14] Proceedings of the NASA Earth Resources Survey Symposium. NASA TM X-58168, Houston, Tex., 1975.
[15] Moik, J. G.: Small Interactive Image Processing System (Smips)—System Description. NASA Goddard Space Flight Center X-650-73-286, 1973.
[16] Moik, J. G.: An Interactive System for Digital Image Analysis. Habilitationschrift (in German), Technical University, Graz, Austria, 1974.
[17] Image Processing System VICAR, Guide to Systems Use. JPL Report 324-IPG/1067, 1968.
[18] Lintz, J.; and Simonett, D. S.: Remote Sensing of Environment. Addison-Wesley, Reading, Mass., 1976.
[19] Nagy, G.: Digital Image Processing Activities in Remote Sensing for Earth Resources, Proc. IEEE, vol. 60, 1972, pp. 1177–1200.
[20] Moik, J. G.: A Data Base Management System for Remote Sensing Data. International Conference on Computer Mapping Software and Data Bases, Harvard University, Cambridge, Mass., July 1978.
[21] Bryant, N. A.; and Zobrist, A. L.: IBIS: A Geographic Information System Based on Digital Image Processing and Image Raster Datatype. Proceedings of a Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Lafayette, Ind., 1976, pp. 1A–1—1A–7.
[22] Bracken, P. A.; Dalton, J. T.; Quann, J. J.; and Billingsley, J. B.: AOIPS—An Interactive Image Processing System. National Computer Conference Proceedings, AFIPS Press, 1978, pp. 159–171.

2. Image Processing Foundations

2.1 Representation of Remotely Sensed Images

Remote sensing derives information by observing and analyzing the spatial, spectral, temporal, and polarization variations of radiation emitted and reflected by the surface of the Earth or by the atmosphere in the optical region of the electromagnetic spectrum. This region extends from X-rays to microwaves and encompasses the ultraviolet, visible, and infrared. Within this broad spectral region, extending from 0.2 to 1,000 μm, the optical techniques of refraction and reflection can be used to focus and redirect radiation.

Remote sensing uses photographic and nonphotographic sensors as imaging devices. Nonphotographic devices are television systems and optical–mechanical scanners. An airborne or spaceborne sensor observes radiation emanating from a scene and modified by the intervening atmosphere. The spectral radiance L of an object at location x, y and at time t has two contributing factors, an emission and a reflectance component [1]

$$L(x, y, \lambda, t, p) = (1 - r(x, y, \lambda, t, p))\, M(\lambda) + r(x, y, \lambda, t, p) i(x, y, \lambda, t) \tag{2.1}$$

The function $r(x, y, \lambda, t, p)$ is the spectral reflectance of the object; $i(x, y, \lambda, t)$ is the spectral irradiance (incident illumination) on the object; and $M(\lambda)$ is the spectral radiant emittance of a blackbody. The parameter p indicates the polarization and λ is the wavelength. L is in general dependent on the solar zenith and on the viewing angle.

In the visible and near-infrared spectrum, where self-emission is negligible and reflected solar energy predominates, the radiance of an object consists of a reflectance and an illumination component. The illumination component is determined by the lighting of the scene, and the reflectance component characterizes the objects or materials in the scene [2]. In the mid- and far-infrared regions, the emission from the surface is dominant. Both reflection and emission must be considered in some microwave experiments [3].

In this text the interest is in analyzing images that are two-dimensional spatial distributions. An image can be represented by a real function of two spatial variables x and y, representing the value of a physical variable at the spatial location (x, y). Therefore, the spatial coordinates x and y of the radiance L are chosen as basic variables, with the spectral, tem-

poral, and polarization variables as parameters. Let $f_i(x, y) = L(x, y, \Delta\lambda_j, t_m, p_n)$ be the spatial distribution for a given spectral band $\Delta\lambda_j$, $j=1, \ldots, P_1$; a given time t_m, $m=1, \ldots, P_2$; and polarization p_n, $n=1, \ldots, P_3$. The $P=P_1+P_2+P_3$ functions $f_i(x, y)$ are combined into the real vector function

$$\mathbf{f}(x, y) = \begin{pmatrix} f_1(x, y) \\ \cdot \\ \cdot \\ \cdot \\ f_P(x, y) \end{pmatrix} \tag{2.2}$$

which will be called multi-image. The measurements in several spectral bands for a given time, ignoring polarization, are called multispectral images. Measurements of different times in a given spectral band are called multitemporal images.

Image functions are defined over a usually rectangular region $R=\{(x, y) \mid 0 \leq x \leq x_m, 0 \leq y \leq y_m\}$. Because energy distributions are nonnegative and bounded, every image function is nonnegative and bounded; i.e.,

$$0 \leq f_i(x, y) \leq B_i \qquad \begin{matrix} i=1, \ldots, P \\ \text{for all } x, y \text{ in } R \end{matrix} \tag{2.3}$$

The orientation of the coordinate system used in this text is shown in figure 2.1, where the x-axis is in the direction of increasing image line numbers. The value of the function f_i at a spatial location (x_0, y_0) is called the gray value of the image component at that point. A P=dimensional

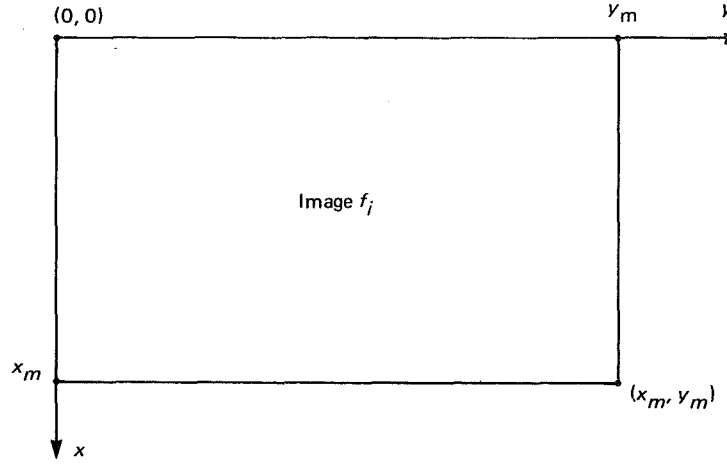


FIGURE 2.1—*Image domain.*

vector $\mathbf{f}(x_0, y_0)$, consisting of the values of $\mathbf{f}$ for a given location (x_0, y_0) is called a multidimensional picture element, or pixel. The range of pixel values is called gray scale, where the lowest value is considered black, and the highest value is considered white. All intermediate values represent shades of gray.

The quality of the information extracted from remotely sensed images is strongly influenced by the spatial, spectral, radiometric, and temporal resolution. Spatial resolution is the resolving power of an instrument needed for the discrimination of observed features. Spectral resolution encompasses the width of the regions in the electromagnetic spectrum that are sensed and the number of channels used. Radiometric resolution can be defined as the sensitivity of the sensor to differences in signal strength. Thus, radiometric resolution defines the number of discernible signal levels. Temporal resolution is defined as the length of the time intervals between measurements. Adequate temporal resolution is important for the identification of dynamically changing processes, such as crop growth, land use, hydrological events, and atmospheric flow.

2.2 Mathematical Preliminaries

Images may be considered as deterministic functions or representatives of random fields. The mathematical tools for image processing are borrowed from the fields of linear system theory, unitary transforms, numerical analysis, and the theory of random fields. This section reviews the required mathematical methods. A complete discussion and proofs can be found in the references.

Image processing involves mathematical operations that require certain analytical behavior of $f_i(x, y)$. It is assumed that the functions representing images are analytically well behaved, i.e., that they are absolutely integrable and have Fourier transforms. The existence of Fourier transforms for functions that are not properly behaved (e.g., constant, impulse, and periodic functions) is guaranteed by assuming the corresponding generalized functions [4, 5]. Existence problems are only of concern for continuous functions. The Fourier transform always exists in the discrete case.

2.2.1 Delta Function and Convolution

The concept of an impulse or a point source of light is very useful for the description and analysis of linear imaging systems. An ideal impulse in the x, y plane is represented by the Dirac delta function $\delta(x, y)$, which is defined as [4, 5]:

$$\int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \delta(x, y)\, dx\, dy = 1 \tag{2.4}$$

$\delta(x, y)=0$ for all (x, y) other than $(0, 0)$, where it is infinite. Useful properties of the delta function for image processing are

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} e^{-2\pi i(ux+vy)}\,du\,dv=\delta(x, y) \tag{2.5}$$

where $i=\sqrt{-1}$ and

$$\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi, \eta)\,\delta(x-\xi, y-\eta)\,d\xi\,d\eta=f(x, y) \tag{2.6}$$

This equation is called the sifting property of the function. The quantities ξ and η are spatial integration variables.

The convolution g of two functions f and h is defined as

$$\begin{aligned} g(x, y) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi, \eta)\,h(x-\xi, y-\eta)\,d\xi\,d\eta \\ &= f(x, y) * h(x, y) \end{aligned} \tag{2.7}$$

where $*$ is the convolution operator. A major application of convolution in image processing is radiometric restoration and enhancement. The autocorrelation of a function $f(x, y)$ is defined as

$$R_{ff}(\xi, \eta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x+\xi, y+\eta)\,f(x, y)\,dx\,dy \tag{2.8}$$

and the crosscorrelation of two functions f and g is

$$R_{fg}(\xi, \eta)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x, y)\,g(x+\xi, y+\eta)\,dx\,dy \tag{2.9}$$

A principal application of crosscorrelation in image processing is image registration where the problem is to find the closest match between two image areas.

2.2.2 Statistical Characterization of Images

For some digital image processing applications an image must be regarded as a sample of a random field rather than a deterministic function. Random fields represent classes of images such as multispectral or multitemporal images of the same or of different scenes. The statistical nature is due to noise and random signal variations in recorded images. The design of some image processing algorithms—for example, classification and image compression techniques—is based on the statistical description of the underlying class of images.

This section provides a brief summary of definitions for random fields that are required later in the text. More information is contained in [6] and [7].

Questions concerning information content and redundancy in images may only be answered on the basis of probability distributions and correlation functions. The entropy H_e of n random variables with probability distribution p_k is defined as

$$H_e = -\sum_{k=1}^{n} p_k \log_2 (p_k) \tag{2.10}$$

The entropy of a probability distribution may be used as a measure of the information content of a symbol (image) chosen from this distribution. The choice of the logarithmic base corresponds to the choice of a unit of information. For the logarithm to the base 2, the unit of information is the bit.

A two-dimensional random field or random process will be denoted $\mathbf{f}(x, y, \omega_i)$ where ω_i is an event and is an element of the set of all events $\Omega = \{\omega_1, \omega_2, \ldots\}$ representing all selections of an image form the given class of images. Each event ω_i is assigned a probability p_i. For a given value of (x, y), $\mathbf{f}(x, y, \omega_i)$ is a random variable, but for a given event ω_i, $\mathbf{f}(x, y, \omega_i)$ is a function over the x, y-plane. Thus, an image $\mathbf{f}(x, y)$ can always be considered as a sample of a random field for a given event. Possible events are the selection of a spectral band, or of a certain time for imaging. The definition of a multi-image in equation 2.2 is an example of a random field. In figure 2.2 the events are selected spectral bands. For a given point (x_0, y_0), $\mathbf{f}(x_0, y_0, \omega_i)$ is a random variable defining a multidimensional picture element (also referred to as a feature vector). For a fixed ω_i, $\mathbf{f}(x, y, \omega_i)$ is a two-dimensional image $\mathrm{f}(x, y)$. A random field representing an image will hence be denoted by $\mathbf{f}(x, y)$.

A random field is completely described by its joint probability density $p_f(z_1, \ldots, z_n; x_1, y_1, \ldots, x_n, y_n)$. In general, higher-order joint probabilities of images are not known. The first-order probability density $p_f(z, x, y)$ can sometimes be modeled on the basis of measurements or properties of the image-generating process. If the statistical properties are independent of the spatial location (x, y), the random field is called homogeneous or stationary [6]. In this case the mean value of $\mathbf{f}$ is defined as

$$\mu_f = E\{\mathbf{f}(x, y)\} = \int_{-\infty}^{\infty} z p_f(z, x, y) dz \tag{2.11}$$

and the autocorrelation function of $\mathbf{f}$ is

$$R_{ff}(\xi, \eta) = E\{\mathbf{f}(x+\xi, y+\eta)\, \mathbf{f}(x, y)\} \tag{2.12}$$

where E is the expectation operator, $\xi = x_1 - x_2$, and $\eta = y_1 - y_2$.

Thus, the mean of a homogeneous random field is constant and the autocorrelation is dependent on the differences of spatial coordinates; i.e.,

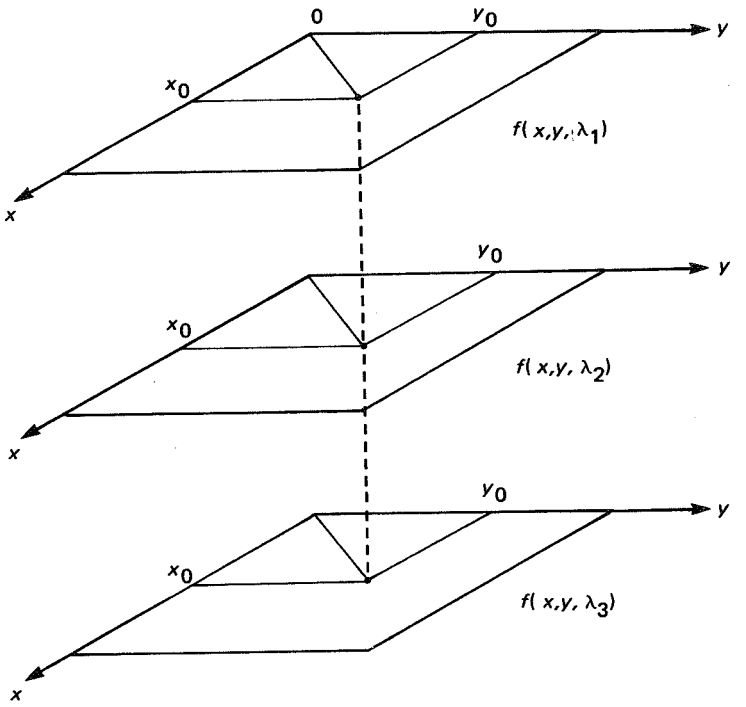


FIGURE 2.2—*Multispectral image (example of a random field).*

the autocorrelation function is position independent. A homogeneous random field **f** representing a class of images may often be assumed to have an autocorrelation function of the form

$$R_{ff}(\xi, \eta) = (R_{ff}(0, 0) - \mu_f^2) e^{-\alpha|\xi| - \beta|\eta|} + \mu_f^2 \tag{2.13}$$

where α and β are positive constants and

$$R_{ff}(0, 0) = E\{\mathbf{f}(x, y)^2\} \tag{2.14}$$

The crosscorrelation function of two real random fields **f** and **g** that are jointly homogeneous is given by

$$R_{fg}(\xi, \eta) = E\{\mathbf{f}(x+\xi, y+\eta)\ \mathbf{g}(x, y)\} \tag{2.15}$$

The covariance function of **f** and **g** is defined by

$$C_{fg}(\xi, \eta) = R_{fg}(\xi, \eta) - \mu_f\, \mu_g \tag{2.16}$$

Two random fields **f** and **g** are called uncorrelated if

$$C_{fg}(\xi, \eta) = 0 \tag{2.17}$$

The Fourier transform (see sec. 2.2.3.1) of the autocorrelation function of a homogeneous random field **f** is the spectral density

$$S_{ff}(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} R_{ff}(\xi, \eta) e^{-2\pi i (\xi u + \eta v)} d\xi \, d\eta \qquad (2.18)$$

The convolution operation, (2.7), is also valid for random fields:

$$\mathbf{g}(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \mathbf{f}(\xi, \eta) \, h(x-\xi, y-\eta) \, d\xi \, d\eta \qquad (2.7)$$

Let $S_{ff}(u, v)$ and $S_{gg}(u, v)$ be the spectral densities of the homogeneous random fields $\mathbf{f}(x, y)$ and $\mathbf{g}(x, y)$, respectively. If **f** has zero mean,

$$S_{gg}(u, v) = S_{ff}(u, v) \mid H(u, v) \mid^2 \qquad (2.19)$$

where $H(u, v)$ is the Fourier transform of $h(x, y)$.

Expressions (2.11) and (2.12) for the mean and the autocorrelation, respectively, of a random field are ensemble averages, each representing a family of equations. In practice, ensemble averages are seldom possible, and the mean and the autocorrelation of a random field are computed as spatial averages. If the autocorrelation functions computed with equation (2.8) for each member of a random field are the same, and if this value is equal to the ensemble average $R_{ff}(\xi, \eta)$, then the autocorrelation function can be obtained from a single image in the field **f** with equation (2.8). Homogeneous random fields for which ensemble and spatial averages are the same are called ergodic. A general discussion of these problems is given in [6].

2.2.3 Unitary Transforms

A unitary transform is a linear transform that expands a function f defined over the region R in the x, y-plane into the sum

$$f(x, y) = \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} F_{\mu\nu} \, \phi_{\mu\nu}(x, y) \qquad (2.20a)$$

The transform coefficients $F_{\mu\nu}$ are given by

$$F_{\mu\nu} = \iint_R f(x, y) \, \phi^*_{\mu\nu}(x, y) \, dx \, dy \qquad (2.20b)$$

where $\phi^*_{\mu\nu}$ may be either real- or complex-valued and $\phi^*_{\mu\nu}$ is the complex conjugate of $\phi_{\mu\nu}$. The expansion (2.20a) is valid if f is square integrable and $\phi_{\mu\nu}(x, y)$ is a complete set of orthonormal functions defined over the same region R of the x, y-plane as $f(x, y)$ [8]. A set of functions $\{\phi_{\mu\nu}(x, y)\}$ is called orthonormal if

$$\iint_R \phi_{\mu\nu}(x, y) \, \phi^*_{\rho\sigma}(x, y) \, dx \, dy = 0 \qquad (2.21)$$

for $\mu \neq \rho$, $\nu \neq \sigma$ and

$$\iint_R |\phi_{\mu\nu}(x, y)|^2 \, dx \, dy = 1 \tag{2.22}$$

It is said to be complete if the mean square error in approximating $f(x, y)$ by (2.20a) approaches zero as the number of terms in (2.20a) approaches infinity.

It can be shown that under the stated conditions

$$\iint_R |f(x, y)|^2 \, dx \, dy = \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} |F_{\mu\nu}|^2 \tag{2.23}$$

This result is known as Parseval's theorem and is a statement of conservation of energy. If only $m \times n$ coefficients are retained, then the coefficients given by equation (2.20b) minimize the mean square error

$$e_{mn} = \iint_R |f(x, y) - \sum_{\mu=0}^{m} \sum_{\nu=0}^{n} F_{\mu\nu} \phi_{\mu\nu}(x, y)|^2 \, dx \, dy \tag{2.24}$$

Complete sets of orthonormal functions are the complex trigonometric functions and the zero-order Bessel functions of the first kind, which define the Fourier and Hankel transforms, respectively.

Unitary transforms can also be applied to a random field representing a class of images. Let $\mathbf{f}(x, y)$ denote a real homogeneous random field with autocorrelation function $R_{ff}(\xi, \eta)$. The expansion (2.20) for the random field $\mathbf{f}(x, y)$ can be expressed as

$$\mathbf{f}(x, y) = \sum_{\mu=0}^{\infty} \sum_{\nu=0}^{\infty} \mathbf{F}_{\mu\nu} \phi_{\mu\nu}(x, y) \tag{2.25a}$$

$$\mathbf{F}_{\mu\nu} = \iint_R \mathbf{f}(x, y) \, \phi^*_{\mu\nu}(x, y) \, dx \, dy \tag{2.25b}$$

The coefficients $\mathbf{F}_{\mu\nu}$ are now random variables having values that depend on the image selected for transformation.

Instead of using a given set of basis functions, the expansion of a random field into a set of orthonormal functions may be adapted to the statistical properties of the class of images under consideration. The expansion is determined such that the coefficients are uncorrelated. Uncorrelated coefficients represent unique image properties. This transform is known as Karhunen–Loéve transform, or transform to principal components. It has the property that for some finite m, n, the mean square error ϵ_{mn} averaged over all images in the random field is a minimum, where

$$\epsilon_{mn} = E\left\{ \iint_R \left| \mathbf{f}(x, y) - \sum_{\mu=0}^{m} \sum_{\nu=0}^{n} \mathbf{F}_{\mu\nu} \phi_{\mu\nu}(x, y) \right|^2 dx \, dy \right\} \tag{2.26}$$

Two-dimensional unitary transforms are used for a number of image processing applications. The design of filters for image restoration and enhancement is facilitated by representing images in the Fourier transform domain. Unitary transforms compress the original image information into a relatively small region of the transform domain. Features for classification are selected from this region, where most of the information resides. Image data compression uses the same property to achieve a bandwidth reduction by discarding or grossly quantizing transform coefficients of low magnitude.

The main advantages of unitary transforms for image processing are that the transform and its inverse can be easily computed, that fast algorithms exist for their computation, and that the information content of images is not modified.

2.2.3.1 Fourier Transform

The two-dimensional Fourier transform of a function $g(x, y)$ is defined by [4]:

$$G(u, v) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(x, y) e^{-2\pi i(ux+vy)}\, dx\, dy \qquad (2.27a)$$

The inverse transform is defined by

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} G(u, v) e^{2\pi i(xu+yv)}\, du\, dv \qquad (2.27b)$$

$G(u, v)$ is a complex-valued function

$$G(u, v) = R(u, v) + iI(u, v) \qquad (2.28)$$

of the spatial frequencies u and v and can be expressed in exponential form as

$$G(u, v) = |G(u, v)|\, e^{i\Phi(u, v)} \qquad (2.29)$$

where

$$|G(u, v)| = \sqrt{R(u, v)^2 + I(u, v)^2} \qquad (2.30)$$

is the magnitude or amplitude of the transform and

$$\Phi(u, v) = \tan^{-1} \frac{I(u, v)}{R(u, v)} \qquad (2.31)$$

is the phase. Figure 2.3*a* shows a block pattern and the magnitude of its Fourier transform.

Let the vectors $\mathbf{z} = (x, y)^T$ and $\mathbf{w} = (u, v)^T$ be the spatial coordinates

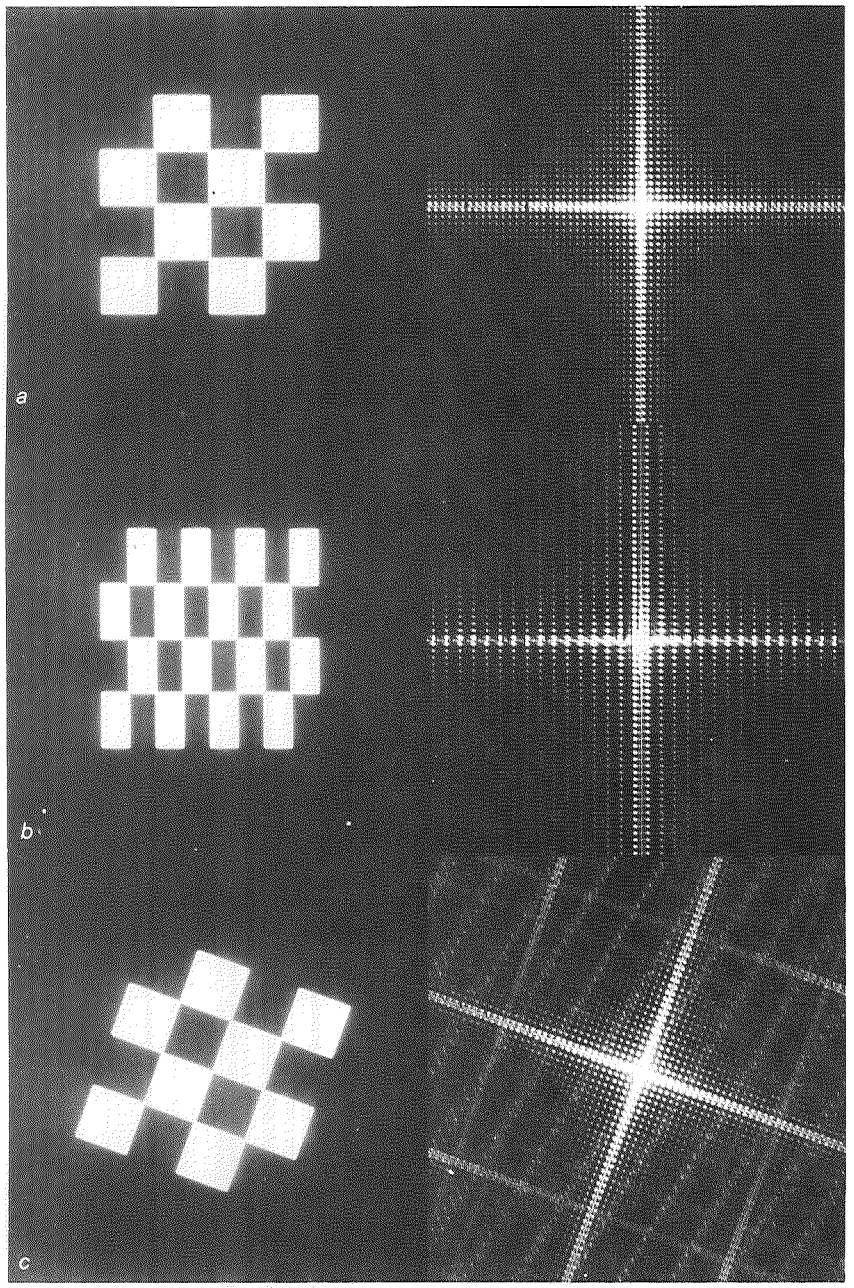


FIGURE 2.3—*Block pattern and magnitude of Fourier transform.*

and the spatial frequencies, respectively. Then, the Fourier transform and its inverse may be written in vector form as

$$G(\mathbf{w}) = \int_{-\infty}^{\infty} g(\mathbf{z}) e^{-2\pi i(\mathbf{w}, \mathbf{z})}\, d\mathbf{z} \tag{2.32a}$$

and

$$g(\mathbf{z}) = \int_{-\infty}^{\infty} G(\mathbf{w}) e^{2\pi i(\mathbf{w}, \mathbf{z})}\, d\mathbf{w} \tag{2.32b}$$

where $(\mathbf{w}, \mathbf{z})$ is the inner product $(\mathbf{w}, \mathbf{z}) = ux + vy$.

Some useful properties of the Fourier transform $\mathcal{F}$ for image processing are—

1. Linearity
2. Behavior under an affine transformation
3. Relationship between a convolution and its Fourier transform
4. Relationship between the crosscorrelation function and its Fourier transform
5. Relationship between the autocorrelation function and its Fourier transform
6. Symmetry

These properties will be described in the following paragraphs.

The Fourier transform is linear; i.e.,

$$\mathcal{F}\{ag_1 + bg_2\} = aG_1 + bG_2 \tag{2.33}$$

An affine transformation is defined as

$$\mathbf{z}' = \mathbf{A}\mathbf{z} + \mathbf{t} \tag{2.34}$$

where

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \text{ and } \mathbf{t} = \begin{pmatrix} t_1 \\ t_2 \end{pmatrix} \tag{2.35}$$

The Fourier transform of $g(\mathbf{z}')$ is

$$\mathcal{F}\{g(\mathbf{A}\mathbf{z} + \mathbf{t})\} = \frac{1}{|J|} e^{-2\pi i(\mathbf{w}, \mathbf{t})}\, G((\mathbf{A}^{-1})^T \mathbf{w}) \tag{2.36}$$

where J is the Jacobian

$$J = \begin{pmatrix} \dfrac{\partial z'_1}{\partial x} & \dfrac{\partial z'_1}{\partial y} \\ \dfrac{\partial z'_2}{\partial x} & \dfrac{\partial z'_2}{\partial y} \end{pmatrix} \tag{2.37}$$

For a shift operation

$$\mathbf{z}' = \mathbf{z} + \mathbf{t} \tag{2.38}$$

the transform matrix **A** is the identity matrix and

$$\mathcal{F}\{g(\mathbf{z}+\mathbf{t})\}=e^{2\pi i(\mathbf{w},\mathbf{t})}\,G(\mathbf{w}) \tag{2.39}$$

i.e., the Fourier transform is shift invariant.

For scaling of the coordinate axes

$$\left.\begin{aligned} x'&=a_{11}\,x\\ y'&=a_{22}\,y\end{aligned}\right\} \tag{2.40}$$

the transformation matrix is

$$\mathbf{A}=\begin{pmatrix} a_{11} & 0\\ 0 & a_{22}\end{pmatrix} \text{ and } \mathbf{t}=0 \tag{2.41}$$

and

$$\mathcal{F}\{g(\mathbf{Az})\}=\frac{1}{|a_{11}\,a_{22}|}\,G\left(\frac{u}{a_{11}},\frac{v}{a_{22}}\right) \tag{2.42}$$

Scaling is illustrated in figure 2.3*b*, where the height of the rectangular blocks in the spatial domain is two times their width. For rotation by the angle ϕ

$$\left.\begin{aligned} x'&=x\cos\phi+y\sin\phi\\ y'&=-x\sin\phi+y\cos\phi\end{aligned}\right\} \tag{2.43}$$

the transformation matrix is

$$\mathbf{A}=\begin{pmatrix} \cos\phi & \sin\phi\\ -\sin\phi & \cos\phi\end{pmatrix} \text{ and } \mathbf{t}=0 \tag{2.44}$$

and

$$\mathcal{F}\{g(\mathbf{Az})\}=G(\mathbf{Aw}) \tag{2.45}$$

The transform is also rotated by the angle ϕ. (See fig. 2.3*c*.)

The convolution, given in equation (2.7), and its Fourier transform are related by:

$$\mathcal{F}\{f(x,y)*h(x,y)\}=F(u,v)H(u,v) \tag{2.46a}$$

$$\mathcal{F}\{f(x,y)h(x,y)\}=F(u,v)*H(u,v) \tag{2.46b}$$

The convolution of two functions in the space domain is equivalent to multiplication in the spatial frequency domain. This relationship is the most important property of the Fourier transform for image processing. It is used to map image restoration and enhancement operations (see chs. 3 and 4) into the frequency domain.

The Fourier transform of the crosscorrelation function, given in equation (2.9), is

$$\mathcal{F}\{R_{fg}(\xi,\eta)\}=F^*(u,v)G(u,v) \tag{2.47}$$

where F^* is the complex conjugate of F.

The Fourier transform of the autocorrelation function, given in equation (2.8), is

$$\mathcal{F}\{R_{ff}(\xi, \eta)\} = |F(u, v)|^2 \tag{2.48}$$

The Fourier transform of the autocorrelation function is called the power spectrum.

If $f(x, y)$ is real,

$$F(-u, -v) = F^*(u, v) \tag{2.49}$$

and the magnitude of the transform is symmetric about the origin. Because image functions are always real, only half of the transform magnitude has to be considered. If $f(x, y)$ is a symmetric function; i.e., if

$$f(x, y) = f(-x, -y) \tag{2.50}$$

the Fourier transform $F(u, v)$ is real.

Equation (2.27*b*) is a representation of $g(x, y)$ as a linear combination of elementary periodic patterns of the form $\exp 2\pi i(ux+vy)$. The function $G(u, v)$ as a weighting factor is a measure of the relative contribution of the elementary pattern, with the spatial frequency components u and v, to the total image. The function $G(u, v)$ is called the frequency spectrum of $g(x, y)$. Edges in pictures introduce high spatial frequencies along a line orthogonal to the edge in the complex frequency plane. Large values of $G(u, v)$ for high spatial frequencies u and v correspond to sharp edges, low values to regions of approximately uniform gray values.

As an example, let

$$f(x, y) = \begin{cases} 1 & \text{for } |x| \leq \frac{1}{2}, |y| \leq \frac{1}{2} \\ 0 & \text{elsewhere} \end{cases} \tag{2.51a}$$

The Fourier transform of $f(x, y)$ is

$$F(u, v) = \frac{\sin \pi u}{\pi u} \frac{\sin \pi v}{\pi v} \tag{2.51b}$$

Equation (2.51a) represents a square aperture. A plot of equation (2.51a) is shown in figure 2.4*a* in three-dimensional perspective. Figure 2.4*b* is a plot of the magnitude, given by equation (2.51b).

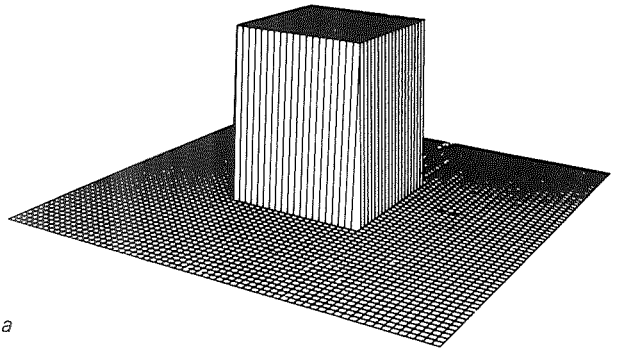
2.2.3.2 Hankel Transform

For circularly symmetric functions $f(x, y) = f([x^2+y^2]^{1/2}) = f(r)$, the transform pair is given by [9]:

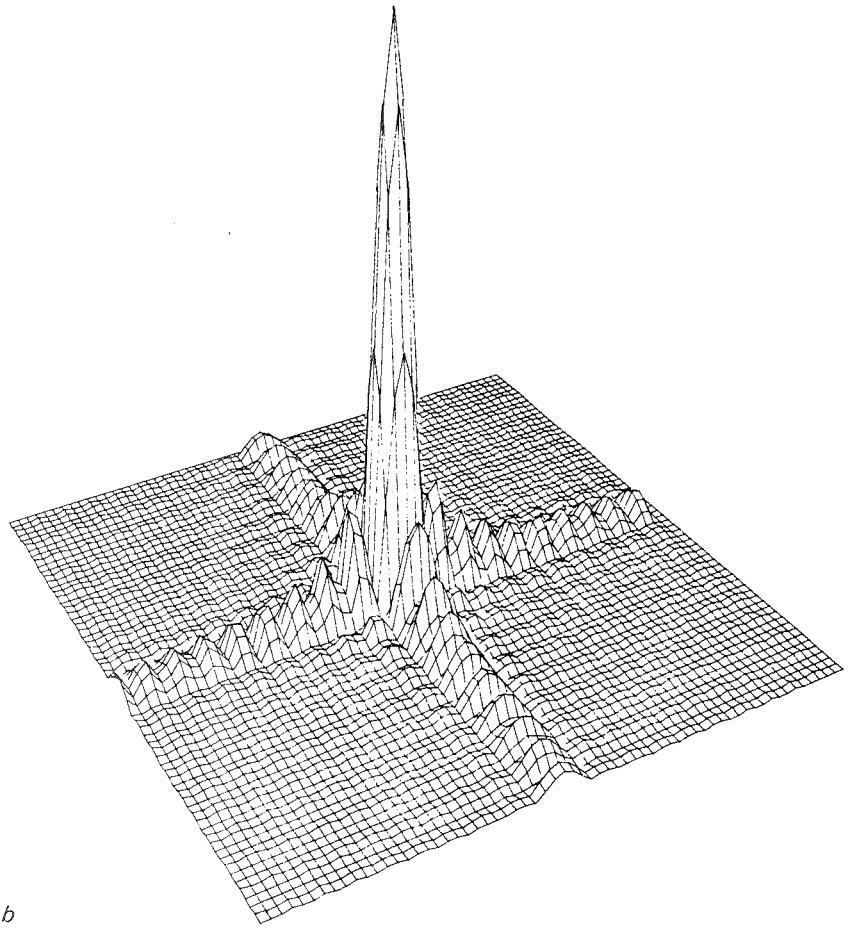
$$F(u, v) = F(\omega) = 2\pi \int_0^\infty rf(r) J_0(2\pi r\omega)\, dr \tag{2.52a}$$

$$f(r) = 2\pi \int_0^\infty \omega F(\omega) J_0(2\pi r\omega)\, d\omega \tag{2.52b}$$

where $\omega = (u^2+v^2)^{1/2}$ is the radial spatial frequency, and $J_0(r)$ is the



a



b

FIGURE 2.4—*Fourier transform pair. (a) Square aperture. (b) Magnitude of Fourier transform.*

zero-order Bessel function of the first kind. The transformation given in equations (2.52a) and (2.52b) is called the Hankel transform and is used in optical image processing where circular apertures can be easily realized.

2.2.3.3 Karhunen–Loéve Transform

The Karhunen–Loéve (K–L) transform does not use a given set of orthonormal functions, but determines the expansion from the statistics of the given class of images. The images are assumed to be representatives of a homogeneous random field whose correlation function is computed from the sample images. For zero-mean random fields with autocorrelation function R the functions $\phi_{\mu\nu}(x, y)$, which yield uncorrelated coefficients $\mathbf{F}_{\mu\nu}$, must satisfy the integral equation [10]:

$$\iint_R R(x-\xi, y-\eta)\phi_{\mu\nu}(x, y)d\xi\, d\eta=\lambda_{\mu\nu}\,\phi_{\mu\nu}(x, y) \qquad (2.53)$$

The possibly complex expansion coefficients $\mathbf{F}_{\mu\nu}$ determined by equation (2.25b) are uncorrelated; i.e.,

$$E\{\mathbf{F}_{\mu\nu}\,\mathbf{F}_{\alpha\beta}{}^*\}=E\{\mathbf{F}_{\mu\nu}\}E\{\mathbf{F}_{\alpha\beta}{}^*\} \qquad (2.54)$$

If a finite number of coefficients are used, an image is optionally approximated with these uncorrelated coefficients in a mean-square-error sense. The Karhunen–Loéve transform is used for image enhancement (ch. 4), feature selection (ch. 8), and image compression (ch. 9).

2.2.4 Description of Linear Systems

Image formation in a linear optical system can be described by a superposition integral [9]. This representation permits the use of linear system theory for the analysis of imaging systems. A linear system $\mathcal{L}$ is an operator that maps a set of input functions into a set of output functions such that

$$\mathcal{L}\{af_1+bf_2\}=a\,\mathcal{L}\{f_1\}+b\,\mathcal{L}\{f_2\} \qquad (2.55)$$

An arbitrary image f can be represented as a sum of point sources

$$f(x, y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi, \eta)\delta(x-\xi, y-\eta)d\xi\, d\eta \qquad (2.6)$$

where δ is the Dirac delta function. The response of a linear system to an input function given in equation 2.6 is

$$\begin{aligned} g(x, y)=\mathcal{L}\{f(x, y)\}&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi, \eta)\mathcal{L}\{\delta(x-\xi, y-\eta)\}d\xi\, d\eta \\ &=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi, \eta)h(x-\xi, y-\eta)d\xi\, d\eta \qquad (2.56) \\ &=f(x, y) * h(x, y) \end{aligned}$$

where $h(x-\xi, y-\eta)$ is the impulse response of the linear space-invariant system $\mathcal{L}$. In other words, the output of $\mathcal{L}$ is found by convolving the input signal with h. Thus, a linear space-invariant system is completely described by its impulse response. In the context of imaging systems, the impulse response h is also called the point spread function (PSF). The PSF is the image of an ideal point source in the object plane.

An alternative representation of a linear space-invariant system is obtained by applying the convolution property (2.46a) to equation (2.56), to yield

$$G(u, v) = F(u, v)H(u, v) \tag{2.57}$$

where $G(u, v)$, $F(u, v)$, and $H(u, v)$ are the Fourier transforms of $g(x, y)$, $f(x, y)$, and $h(x, y)$ respectively. $H(u, v)$, the Fourier transform of the PSF $h(x, y)$, is called the optical transfer function (OTF) of the linear space-invariant imaging system. The OTF, which is generally complex, can be expressed in exponential form as

$$H(u, v) = M(u, v)e^{i\Phi(u, v)} \tag{2.58}$$

The amplitude $M(u, v)$ and phase $\Phi(u, v)$ are called the modulation transfer function (MTF) and phase transfer function (PTF), respectively.

2.2.5 Filtering

Filtering is a basic image processing operation used in radiometric restoration and enhancement of images. A linear filter is a linear space-invariant system that modifies the spatial frequency characteristics of an image. Because the effect of any linear space-invariant system can be represented by a convolution, linear filtering may be described by

$$g_f(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} g(\xi, \eta)h(x-\xi, y-\eta)\,d\xi\,d\eta \tag{2.59}$$

in the spatial domain, or by

$$G_f(u, v) = G(u, v)H(u, v) \tag{2.60}$$

in the frequency domain. (See the Fourier transform property given in eq. 2.46a.) The quantity g is the recorded image with Fourier transform G, h is the impulse response of the filter with Fourier transform H, and g_f is the filtered image with Fourier transform G_f. The variable H is called the filter transfer function. Only nonrecursive filters will be considered here. Because the entire digital image is recorded, ideal filters can be realized numerically, and the impulse response $h(x, y)$ may be symmetric. This symmetry implies a purely real transfer function $H(u, v)$, i.e., a phaseless filter.

Filters are conceptually easier to specify and apply in the frequency domain rather than in the spatial domain, because their representation in the frequency domain is simpler and convolution is replaced by multiplication [11, 12]. Filters may be characterized in the frequency domain by the shape of their transfer function $H(u, v)$. Circularly symmetric filters represented in polar coordinates will be considered where the radial spatial frequency coordinate is $\omega = (u^2 + v^2)^{1/2}$.

An ideal low-pass filter is defined by

$$H(u, v) = H(\omega) = \begin{cases} 1 \text{ for } \omega \leq \omega_c \\ 0 \text{ for } \omega > \omega_c \end{cases} \tag{2.61}$$

It suppresses all frequencies above the cutoff frequency ω_c. An ideal high-pass filter is defined by

$$H(\omega) = \begin{cases} 1 \text{ for } \omega \geq \omega_c \\ 0 \text{ for } \omega < \omega_c \end{cases} \tag{2.62}$$

It suppresses all frequencies below the cutoff frequency. A band-pass filter is a combination of low- and high-pass filters; it suppresses all spatial frequencies outside its pass band. A notch filter is the inverse of a band-pass filter; it suppresses all frequencies in a specified band.

An exponential filter is defined by

$$H(\omega) = \begin{cases} ce^{a\omega^2} \text{ for } \omega \leq \omega_c \\ H_{\max} \text{ for } \omega > \omega_c \end{cases} \tag{2.63}$$

A Gaussian filter is defined by

$$H(\omega) = ce^{-a\omega^2} \tag{2.64}$$

Because of the convolution property, linear filtering may be performed either in the frequency domain with equation (2.60) or in the spatial domain with equation (2.59). For frequency-domain filtering, the multiplication of $G(u, v)$ with an ideal filter causes discontinuities that result in oscillations of g_f in the spatial domain due to the Gibbs phenomenon [13]. These oscillations become apparent as ringing in the filtered image. Consequently, filters have to be designed so that ringing is minimized.

Spatial-domain filters are usually designed in the frequency domain by specifying the filter transfer functions $H(u, 0)$ and $H(0, v)$ for the main frequency axes. Under the assumption that curves of constant $H(u, v)$ are ellipses, the transfer function $H(u, v)$ may be computed. The filter response is given by the inverse Fourier transform

$$h(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} H(u, v) e^{2\pi i(ux + vy)} \, du \, dv \tag{2.65}$$

which is of infinite extent [12]. In reality, a finite-extent filter response h_t is used, obtained by truncating $h(x, y)$ with a window $w(x, y)$:

$$h_t(x, y) = h(x, y)w(x, y) \tag{2.66}$$

Because of the property given by equation (2.46b), the actual filter transfer function is then the convolution of the ideal transfer function with the Fourier transform of the window function

$$H_z(u, v) = H(u, v) * W(u, v) \tag{2.67}$$

Spatial-domain truncation also causes overshooting or ringing in the neighborhood of edges in the filtered image. To reduce ringing, the window function has to be chosen so that $H_t(u, v)$ is close to $H(u, v)$ and that the ripples of $H_t(u, v)$ in the neighborhood of discontinuities of $H(u, v)$ are small. A window or apodizing function should only affect the border region of an image [14]. For circularly symmetric filters one-dimensional windows $w(x, y) = w(x^2+y^2)^{1/2} = w(r)$ can be used [15]. For a constant window function

$$w(r) = \begin{cases} 1 \text{ for } |r| \leq r_0 \\ 0 \text{ for } |r| > r_0 \end{cases} \tag{2.68}$$

where r_0 is the width of the filter. The Fourier transform is

$$W(\omega) = \frac{\sin 2\pi r_0 \omega}{2\pi r_0 \omega} \tag{2.69}$$

whose first side-lobe peak is about 23 percent of the peak at $\omega = 0$. The variable ω is the radial spatial frequency. For a triangular window

$$w(r) = \begin{cases} 1 - \dfrac{|r|}{r_0} \text{ for } |r| \leq r_0 \\ 0 \quad \text{for } |r| > r_0 \end{cases} \tag{2.70}$$

the Fourier transform is

$$W(\omega) = \left(\frac{\sin \dfrac{\pi\omega}{2}}{\dfrac{\pi\omega}{2}} \right)^2 \tag{2.71}$$

whose first side-lobe peak is about 4 percent of the main peak.

For the Hanning window

$$w(r) = 0.54 + 0.46 |r| \tag{2.72}$$

the absolute value of the largest side lobe in $W(\omega)$ is less than 1 percent of the main peak. Another frequently used window is the raised cosine bell, or Hanning window.

$$w(r)=\begin{cases}0.5\left(1-\cos\dfrac{\pi r}{pr_0}\right) & 0\leq|r|\leq pr_0\\ 1 & pr_0\leq|r|\leq(1-p)r_0\\ 0.5\left(1-\cos\dfrac{\pi(r_0-r)}{pr_0}\right) & r_0(1-p)\leq|r|<r_0\end{cases} \quad (2.73)$$

where p is the fraction of the filter width over which the window is applied [14].

Filter design consists of the following steps:

1. Specify the ideal filter transfer function $H(u, v)$ in frequency domain.
2. Compute filter response $h(x, y)$ as inverse Fourier transform of $H(u, v)$.
3. Multiply filter h by a proper window function w.
4. Fourier transform the apodized filter h_t and evaluate the resulting transfer function H_t.

The following examples are only intended to illustrate the effects of ideal filters. Because of their undesirable side effects (ringing), ideal filters are not used for filtering applications. The image used in the examples was generated by digitizing a NASA slide to 512×512 samples on the Atmospheric and Oceanographic Image Processing System [16].

Figure 2.5 shows the image and the magnitude of its Fourier transform. It is obvious that most of the image energy is concentrated in a region near the origin. The results of applying circularly symmetric filters to the image in figure 2.5 are shown in figures 2.6 to 2.8.

Figure 2.6 is filtered with an ideal low-pass filter with cutoff frequency ω_c where $\omega_c=(u_c^2+v_c^2)^{1/2}=0.2$. The blurring indicates that most of the edge information in the image is contained in a region above ω_c. Severe ringing effects are visible.

Figure 2.7 shows the result of applying an ideal high-pass filter with cutoff frequency $\omega_c=0.1$ to the image in figure 2.5. Figure 2.8 is filtered with a Gaussian filter with a width of $w_c=0.12$ at half maximum. The magnitude of the Fourier transform shows that most of the edge information has been removed, and severe blurring results. In section 2.2.3.1 it was shown that the Fourier coefficients in a given direction represent features orthogonal to that direction. Wedge filtering can be used to extract the information representing features in a given direction. Figure 2.9 shows the result of applying a wedge filter of width 30° in a direction of 50°.

Although the magnitude $|G(u, v)|$ of the Fourier transform of an image $g(x, y)$ represents the sharpness of edges, it is the phase $\Phi(u, v)$ that contains the position information of objects in the image. Figure 2.10*b*

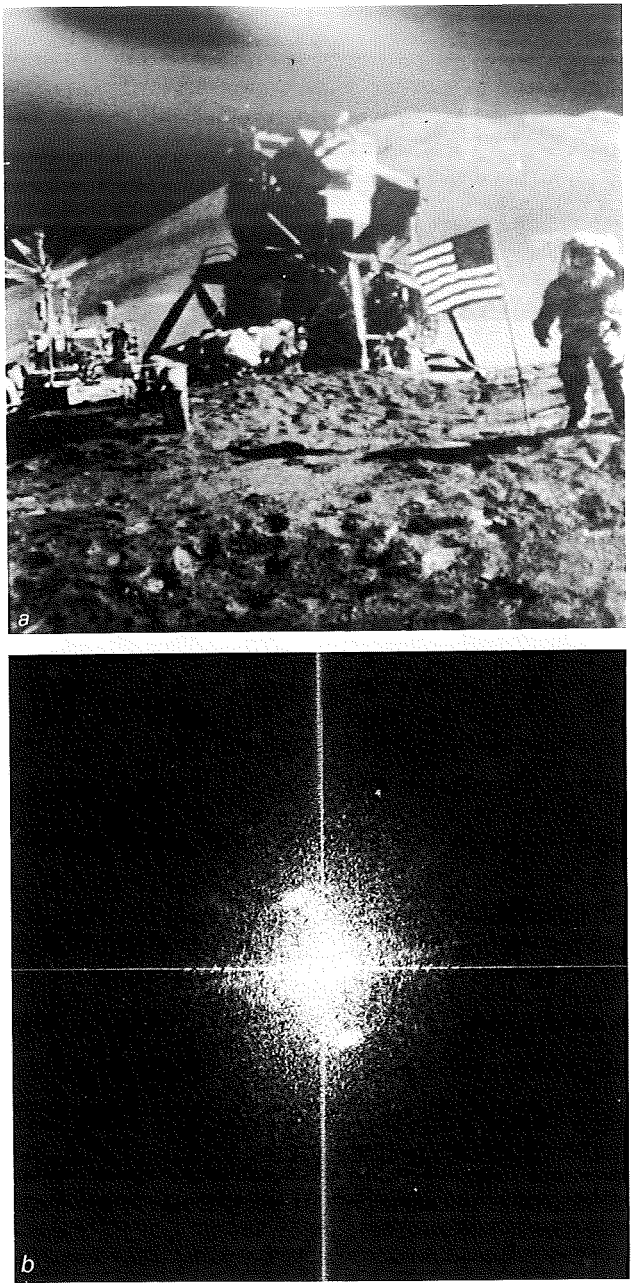


FIGURE 2.5—*Moon image (a) and magnitude of Fourier transform (b).*

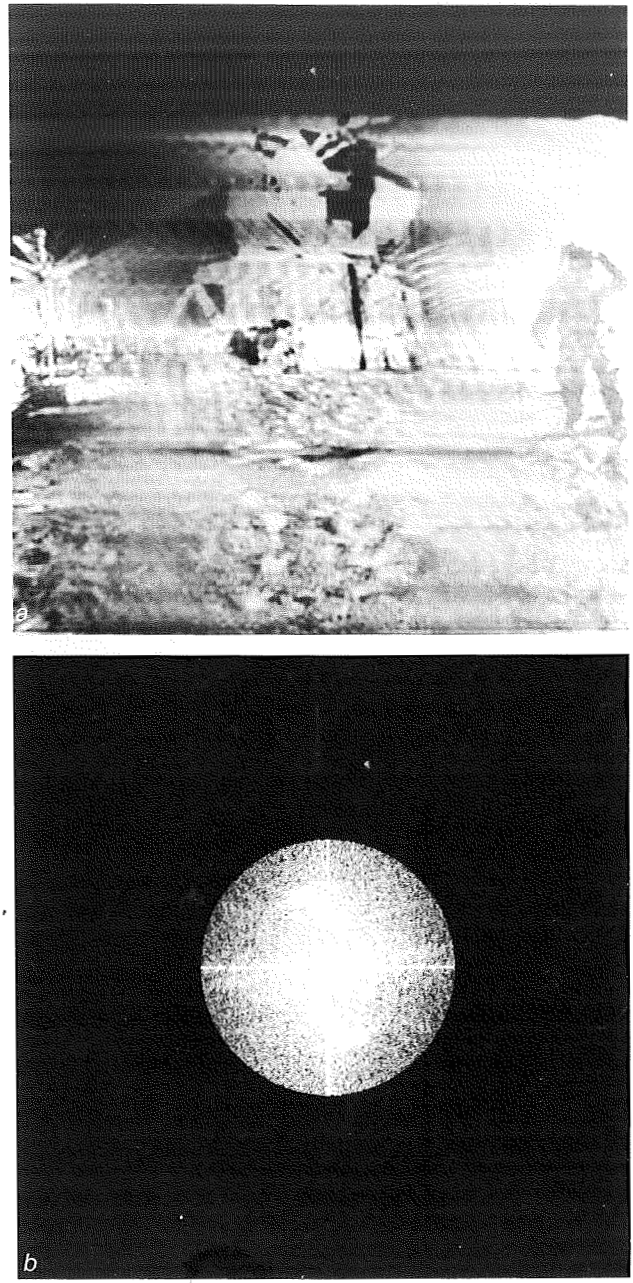


FIGURE 2.6—*Image resulting from low-pass filtering (a) and magnitude of Fourier transform (b).*

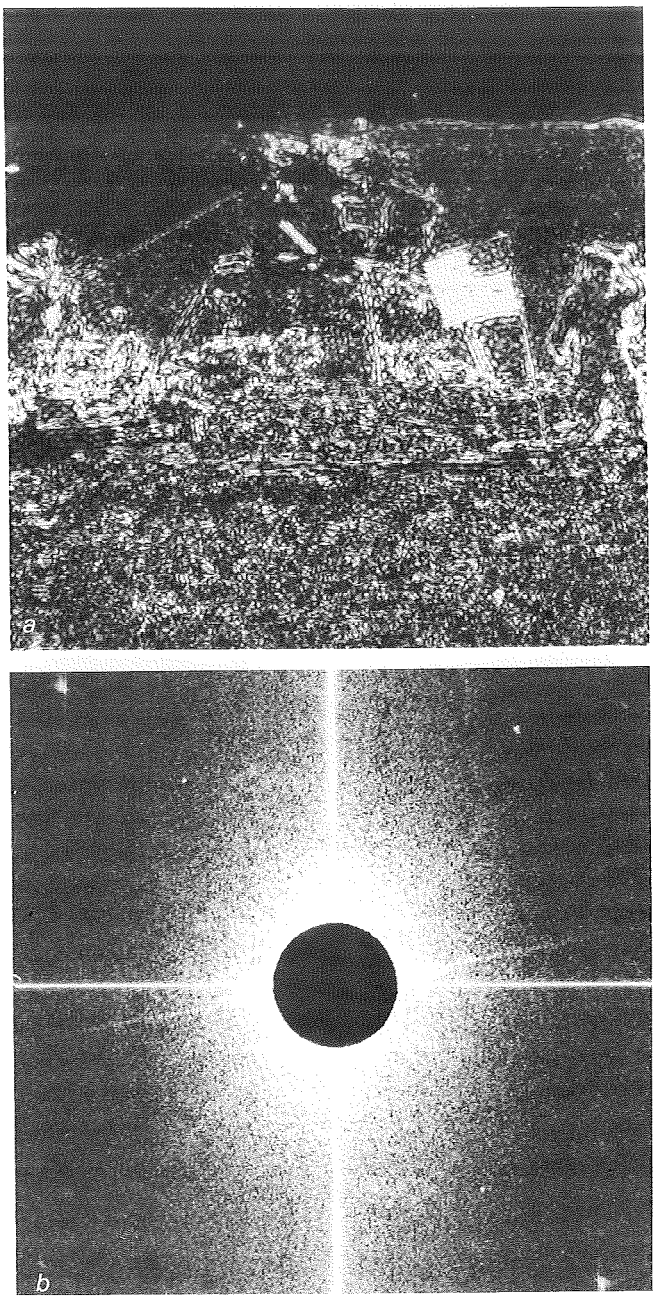


FIGURE 2.7—*Image resulting from high-pass filtering (a) and magnitude of Fourier transform (b).*

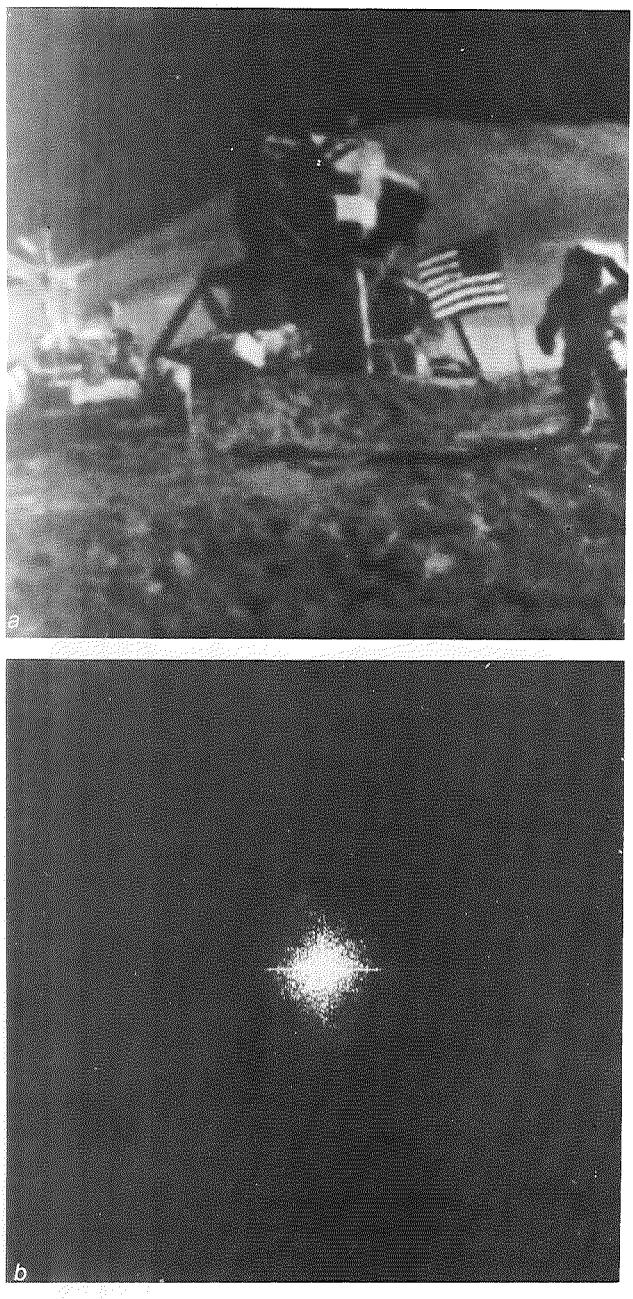


FIGURE 2.8—*Image resulting from Gaussian filtering (a) and magnitude of Fourier transform (b).*

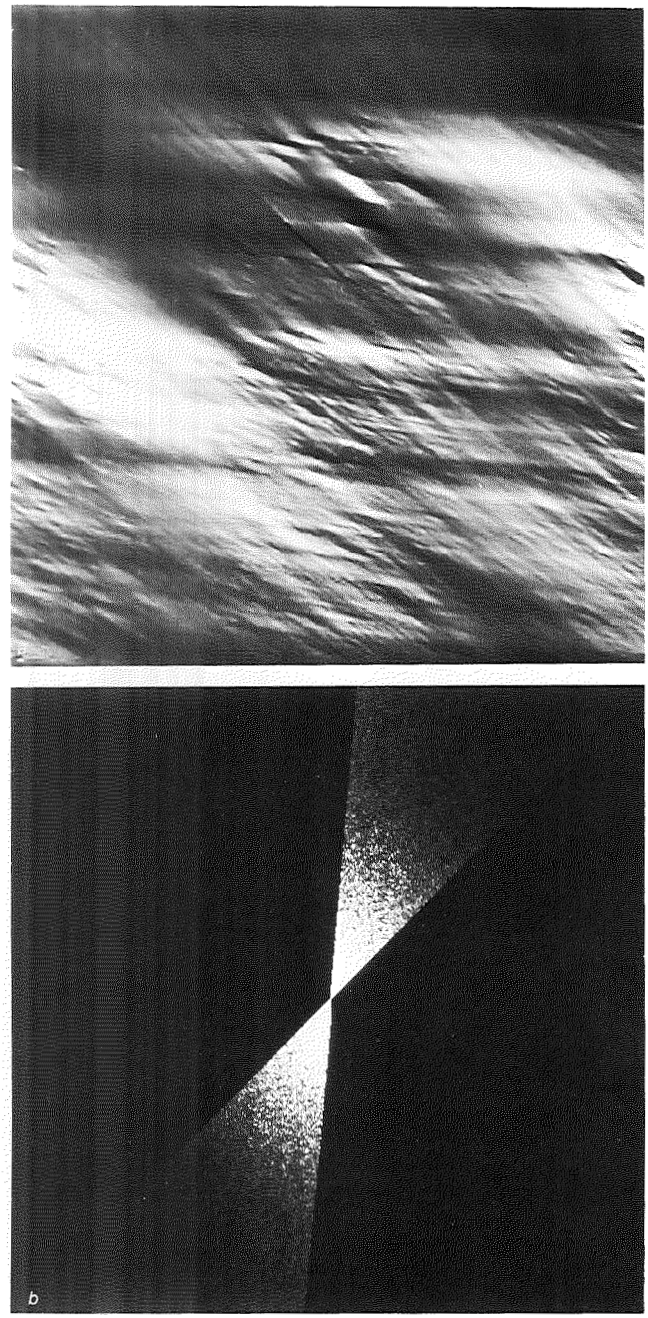


FIGURE 2.9—*Image resulting from wedge filtering (a) and magnitude of Fourier transform (b).*

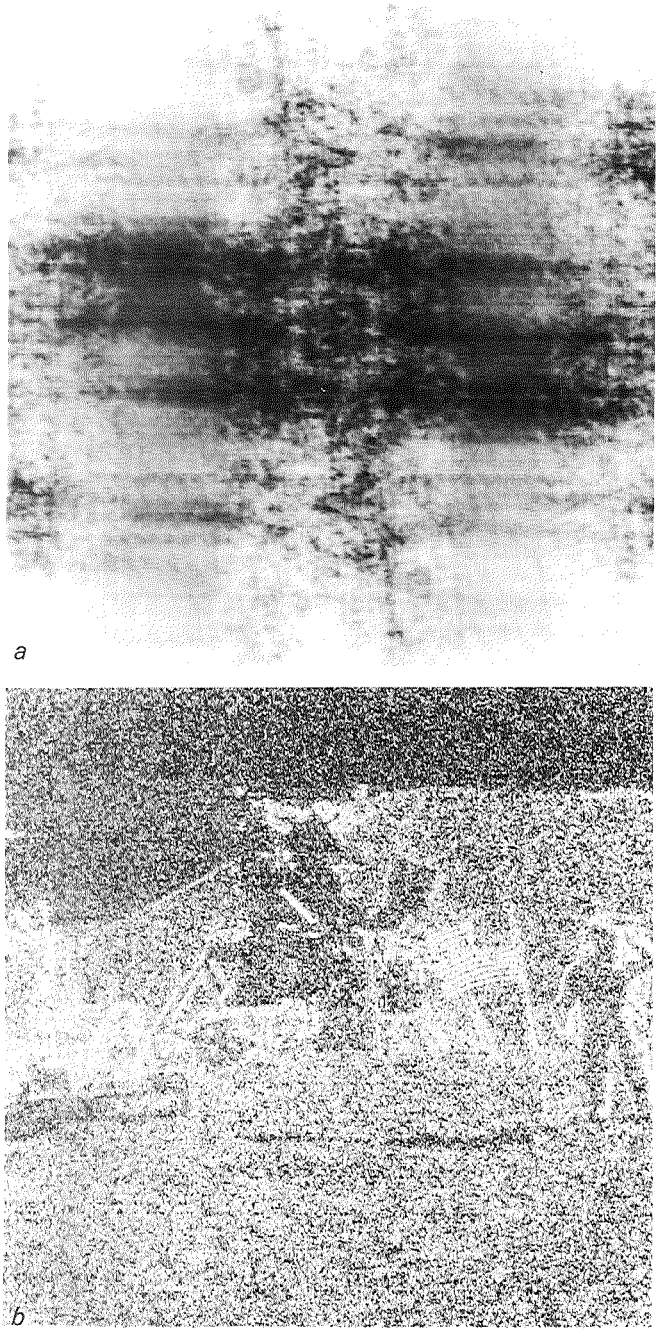


FIGURE 2.10—*Magnitude and phase images of Fourier transform of figure 2.5a. (a) Image reconstructed from magnitude. (b) Image reconstructed from phase component.*

shows the phase image of the Moon image in figure 2.5. This phase image was obtained by setting $|G(u, v)| = 1$ in the u, v plane. Figure 2.10*a* shows the corresponding magnitude image reconstructed from the Fourier transform after setting the phase $\Phi(u, v) = 0$ in the u, v plane.

For certain nonlinear systems, generalized linear filtering or homomorphic filtering may be used [17]. A nonlinear transformation A maps the system into a new space in which linear filtering may be performed. The filtered output is transformed back into the original space with the inverse transformation A^{-1}. For multiplied signals, as in the image model given in equation (2.1), the logarithm maps multiplication into addition. Homomorphic filtering has been very successfully applied in image processing [2]. A block diagram of homomorphic filtering is shown in figure 2.11. Figure 2.12 shows an original image and the result of homomorphic filtering with a high-pass filter that suppresses the illumination component while enhancing the reflectance part of an image.

2.3 Image Formation and Recording

Imaging systems transform variations in object radiant energy into an image or an electrical signal from which an image can be reconstructed. Two basic systems are employed to accomplish this function, frame cameras and scanning cameras. Frame cameras, e.g., film and storage television cameras, sense all picture elements of an object simultaneously. Scanning cameras, such as conventional television cameras and optical–mechanical scanners, sense the picture elements time sequentially. Images are formed by an optical system that projects the radiation onto a photosensor in the image plane. The photosensor converts the radiation into a latent image on photographic film or into an electric signal that is amplified for analog transmission and amplified, sampled, and quantized for digital transmission.

Images are always degraded to some extent because of atmospheric effects and the characteristics of the sensing and recording system. These degradations may be grouped into radiometric and geometric distortions. Radiometric degradations arise from blurring effects of the imaging system, nonlinear amplitude response, vignetting and shading, transmission noise, atmospheric interference (scattering, attenuation, and haze),

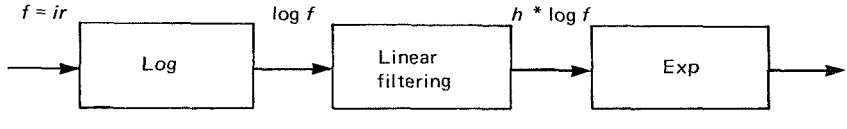


FIGURE 2.11—*Block diagram of homomorphic filtering.*

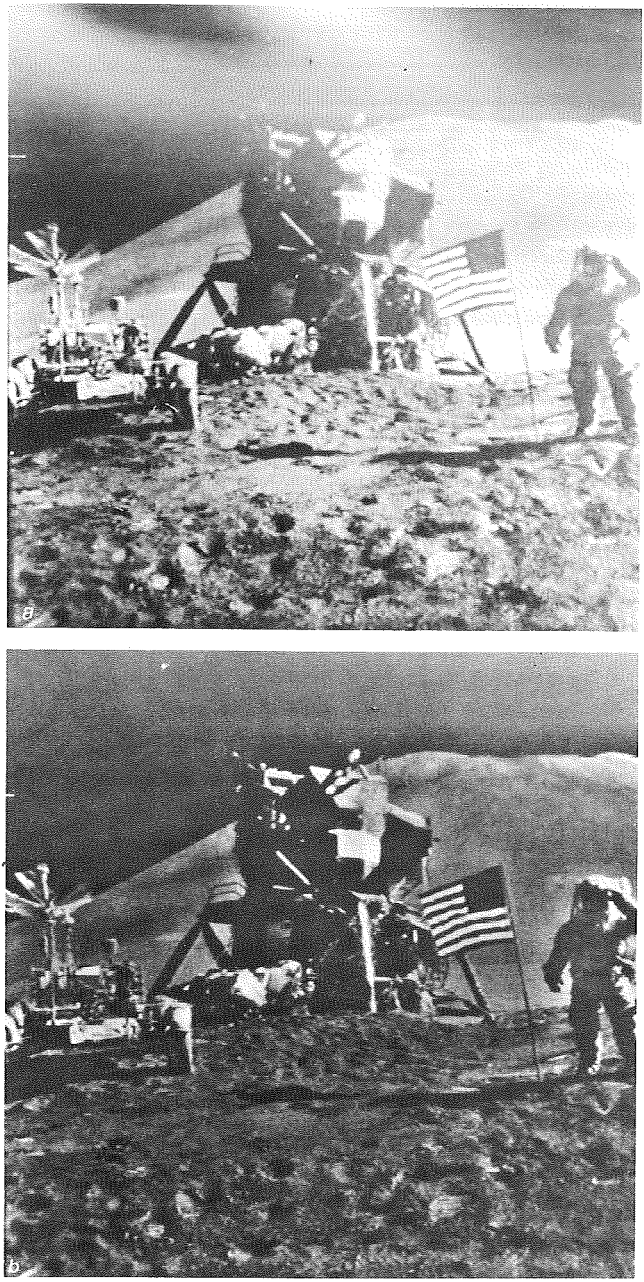


FIGURE 2.12—*Homomorphic filtering. (a) Original image. (b) Result of homomorphic filtering with a high-pass filter.*

variable surface illumination (differences in terrain slope and orientation), and change of terrain radiance with viewing angle.

Geometric distortions can be categorized into sensor-related distortions, such as aberrations in the optical system; nonlinearities and noise in the scan-deflection system; sensor-platform-related distortions caused by changes in the attitude and altitude of the sensor; and object-related distortions caused by Earth rotation, Earth curvature, and terrain relief. Perspective distortions are dependent on the type of camera. Although ideal images from frame cameras are perspective projections of the Earth scene, images from scanning cameras exhibit a variable distortion because of the combined motions of sensor and object during the imaging time. Knowledge of the processes by which images are formed is essential for image processing. The effects of the sensing and recording device on images must be understood to correct for the distortions.

Image formation and recording can be mathematically described by a transformation T that maps an object $f_i(x', y')$ from the object plane with coordinate system (x', y') into the recorded image $g_i(x, y)$ with coordinate system (x, y),

$$g_i(x, y) = T_i\{f_i(x', y')\} \qquad i = 1, \ldots, P \tag{2.74}$$

where T_i represents the image degradations for the ith component of the multi-image. To recover the original information from the recorded observations, the nature of the transformation T_i must be determined, followed by the inverse transformation T_i^{-1} on the image $g_i(x, y)$. The following discussion will refer to component images and the index i will be omitted.

The mathematical treatment is facilitated by separating the degradations into geometric distortions T_G and radiometric degradations T_R

$$g(x, y) = T_G T_R\{f(x', y')\} \tag{2.75}$$

Geometric distortions affect only the position rather than the magnitude of the gray values. Thus, T_G is a coordinate transformation, which is given by

$$\left.\begin{aligned} x' &= p(x, y) \\ y' &= q(x, y) \end{aligned}\right\} \tag{2.76}$$

The radiometric degradation T_R represents the effects of atmospheric transfer, image formation, sensing, and recording the image intensity distribution. The influence of the atmosphere on the object radiant energy f is determined by attenuation, scattering, and emission. A fraction τ $(0 < \tau < 1)$ of the emitted and reflected object radiance is transmitted to the sensor. The radiance scattered or emitted by the atmosphere into the sensor's field of view is B and is sometimes called path radiance [3]. Thus,

the object distribution $f(x, y)$ is modified by the atmosphere into the apparent object radiance

$$f_i^*(x, y) = \tau_i(x, y) f_i(x, y) + B_i(x, y) \qquad i = 1, \ldots, P \quad (2.77)$$

where τ_i is the spectral transmittance of the atmosphere. The path radiance B_i, consisting of a scattering and an emission component, limits the amount of information that can be extracted from the measured radiation.

Image formation is the transformation of the apparent object radiant energy $f^*(x', y')$ in the object plane into image radiant energy $g_0(x', y')$ in the image plane. Image formation in a linear optical system under the assumption of an ideal aberration-free lens can be described by [9, 18]

$$g_0(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_0(x', y', \xi, \eta) f^*(\xi, \eta) d\xi \, d\eta \quad (2.78)$$

This equation expresses the fact that the radiant energy distribution in the image plane is the superposition of infinitesimal contributions due to all object point contributions. The function h_0 is the PSF of the optical system. It determines the radiant energy distribution in the image plane due to a point source of radiant energy located in the object plane. The PSF $h_0(x', y', \xi, \eta)$ describes a space-variant imaging system, because h_0 varies with the position in both image and object plane. If the imaging system acts uniformly across image and object planes, the PSF is independent of position. For such a space-invariant system the image formation equation (2.78) becomes a convolution:

$$g_0(x', y') = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h_0(x' - \xi, y' - \eta) \, f^*(\xi, \eta) d\xi \, d\eta \quad (2.79)$$

The image radiant energy g_0 is sensed and recorded by a sensor. Image sensing and recording can be performed by photochemical and electro-optical systems. Photochemical technology combines detection and recording into photographic film. The nonlinear film characteristic relates the incident light to optical density. The recorded image g represents the local density of photoactivated silver grains. In electro-optical systems, such as vidicons and optical–mechanical scanners, image detection and recording are separate processes. The images are recorded as electrical signals that are well suited for conversion to a digital representation. In these systems the sensor output is a function of the incident light intensity. The sensing and recording induces noise, which is, in general, signal dependent. To facilitate mathematical treatment, noise is assumed to be signal independent and additive.

Noise may be grouped into coherent and random noise. Random noise is denoted by n_r and is assumed to be a homogeneous random process that is uncorrelated with the signal. The effects of sensing and recording are

represented by a nonlinear operator. Often a linear approximation to the nonlinear sensor response [18], as well as to the influence of the atmosphere, is justified. Under this assumption the radiometric transformation T_R, representing atmospheric transfer, optical image formation, sensing and recording for space-invariant systems yields a radiometrically degraded image g_R given by

$$\left.\begin{aligned} g(x', y') &= \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x'-\xi, y'-\eta) f(\xi, \eta)\, d\xi\, d\eta + n_r(x', y') \\ &= T_R f(x', y') \end{aligned}\right\} \tag{2.80}$$

The PSF h of the linearized imaging system is then a combination of the transfer characteristics of the atmosphere, of the optical system, of the photosensor aperture, and of the electronic system. Practically, h is zero outside a region $R_h = \{(x, y) \mid 0 \leq x \leq x_h,\ 0 \leq y \leq y_h\}$. This formulation of the imaging process is identical for frame and scanning cameras. The optical–mechanical line scanning process, however, is a function of time. The formulation of the imaging process given in equation (2.80) implies that the convolution of the object radiance distribution $f(x', y')$ with the system PSF $h(x', y')$ be performed for each picture element. If neither object radiance distribution nor camera response varies with time during imaging, the convolution can be performed for the entire image.

Equation (2.80) represents only the spatial characteristics of object and imaging system. Inclusion of the spectral characteristics would require integration of equation (2.80) over wavelength. The formulation (2.80) is convenient if the spectral characteristics of object and camera are approximately constant over the measured spectral band. This assumption is justified for narrow spectral bands, as in the simplified image representation (eq. 2.1).

A frame camera such as the Return Beam Vidicon (RBV) camera on the near polar-orbiting Landsat 1 and 2 spacecrafts operates by shuttering three independent cameras simultaneously, each sensing a different spectral band. On Landsat 3, two panchromatic cameras are used whose shutters are opened sequentially, producing two side-by-side images rather than three overlapping images of the same scene. The viewed ground scene is stored on the photosensitive surface of the camera tube, and after shuttering, the image is scanned by an electron beam to produce a video signal. To produce overlapping images along the direction of spacecraft motion, the cameras are shuttered every 25 s. The video bandwidth during readout is 3.2 MHz [19].

A scanning camera like the MSS on the Landsat spacecrafts uses an oscillating mirror to continuously scan lines perpendicular to the spacecraft velocity. The image lines are scanned simultaneously in each spectral band for each mirror sweep. The spacecraft motion provides the along-track progression of the scan lines to form an image. The optical energy

from the mirror is sensed simultaneously by an array of six detectors per spectral band. The detector outputs are digitized (see sec. 2.5) and formatted into a continuous data stream of 15 megabits per second. The spatial resolution is approximately 80 m.

Meteorological remote sensing uses spin-stabilized geostationary satellites such as the Synchronous Meteorological Satellite (SMS), which are placed in a synchronous orbit approximately 22,300 miles above the Earth's surface. At this altitude, the satellite orbital angular velocity matches the rotational angular velocity of the Earth so that its position relative to features on the Earth's surface is fixed for a truly equatorial orbit. The satellite is rotated about its axis, which is perpendicular to the orbit plane, at a constant rate so that image lines in an equatorial direction are scanned. An optical system focuses the received radiation onto a detector. To form an image, a stepping mirror deflects the radiation at a different angle on each rotation. The electrical detector output is digitized for transmission to the ground. The SMS has two channels, a visible (0.55–0.7 μm) with a 0.9-km spatial resolution, and a thermal infrared (10.5–12.6 μm) channel with an 8-km spatial resolution. A rapid-scan mode permits increasing the temporal resolution to 3 min.

The complete linearized model for the imaging process is

$$g(x, y) = T_G T_R\{f(x', y')\} + n_s(x, y) \tag{2.81}$$

The radiometrically degraded image $g_R(x', y')$ is geometrically distorted by the operator T_G and coherent noise n_s is added to obtain the final recorded image $g(x, y)$. A block diagram of this model is shown in figure 2.13.

2.4 Degradations

The sources of degradations in imaging systems represented in the image formation model, equation (2.81), can be grouped into several categories.

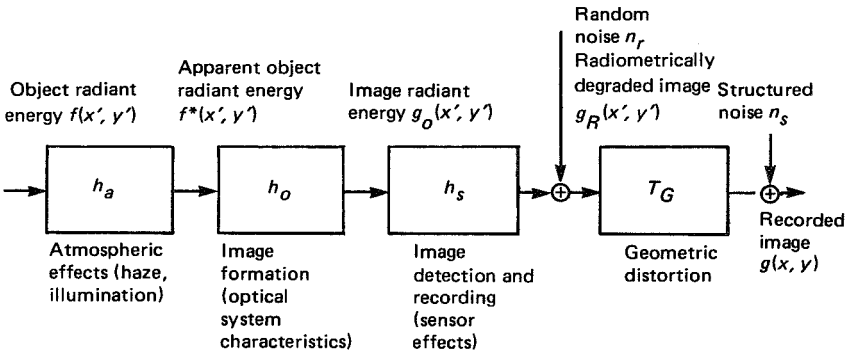


FIGURE 2.13—*Linearized imaging model.*

These categories give rise to corresponding classes of image processing techniques that are used to correct for the degradations. A perfect imaging system would cause no geometric distortions; i.e., $T_G=I$ where I is the identity transformation, or

$$\left.\begin{aligned} x'=x=p(x,y)\\ y'=y=q(x,y)\end{aligned}\right\} \tag{2.82}$$

It would induce no noise:

$$\left.\begin{aligned} n_r(x,y)=0\\ n_s(x,y)=0\end{aligned}\right\} \tag{2.83}$$

and it would have an ideal PSF:

$$h(x,y)=\delta(x,y) \tag{2.84}$$

Thus, the recorded image, given by

$$g(x,y)=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi,\eta)\delta(x-\xi,y-\eta)\,d\xi\,d\eta=f(x,y) \tag{2.85}$$

is identical to the object f because of equation (2.6).

Degradation categories that may be distinguished are discussed in sections 2.4.1 to 2.4.4.

2.4.1 Geometric Distortion

In the absence of spatial and point degradation and noise, and with geometric distortions given in equation (2.76), equation (2.81) becomes

$$\left.\begin{aligned} g(x,y)&=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(\xi,\eta)\delta[p(x,y)-\xi,q(x,y)-\eta]d\xi\,d\eta\\ &=f[p(x,y),q(x,y)]\end{aligned}\right\} \tag{2.86}$$

This equation describes an imaging system that introduces only a distortion due to a coordinate change.

Geometric distortions can be caused by:

1. *Instrument errors*—Examples are distortions in the optical system, scan nonlinearities, scan-length variations, and nonuniform sampling rate.
2. *Panoramic distortion or foreshortening*—This error is caused by scanners using rotating mirrors with constant angular velocity. The velocity $\dot{x}$ of the scanning aperture over the Earth's surface is given by (see fig. 2.14*a*):

$$\dot{x}=\frac{a\dot{\theta}}{\cos^2\theta} \tag{2.87}$$

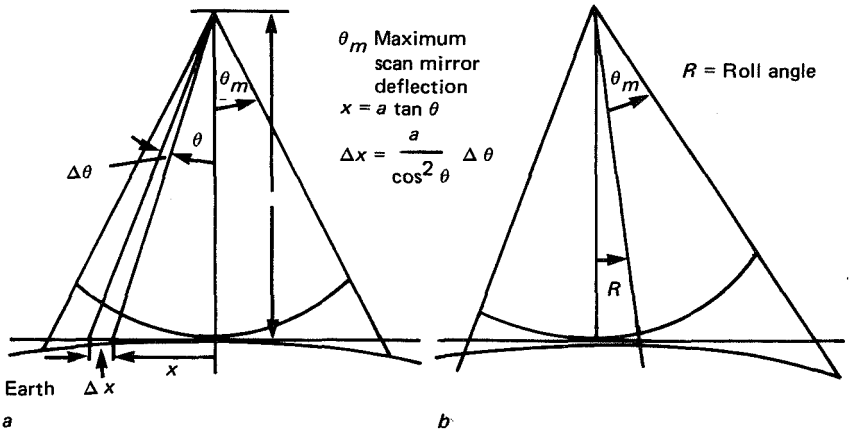


FIGURE 2.14—*Panoramic distortion and roll effect. (a) Panoramic distortion. (b) Effect of roll.*

where a is the altitude of the sensor and θ is the angle of the scan-mirror deflection. Although the scanning aperture velocity is nonlinear, the produced image is recorded with constant velocity. Because of this difference in scanning and recording speed, the distance between sample centers and the sample size in the scan direction are functions of the mirror deflection θ (fig. 2.14*a*). The effect is a scale distortion that increases with the deflection of the mirror from the vertical. For example, the maximum mirror deflection for the Landsat MSS is 5.78°, resulting in a cumulative distortion of about 11 pixels.

3. *Earth rotation*—Significant Earth rotation during the time required to scan a frame causes a skew distortion that varies with latitude. For 40° north latitude the skew angle for Landsat MSS images is about 3°, resulting in a shift of 122 pixels between the top and bottom lines of a frame.
4. *Attitude changes that occur during the time to scan a frame*—These changes are yaw, pitch, and roll. Yaw is the rotation of the air or spacecraft about the local zenith vector (pointed toward the center of the Earth). Yaw causes rotation or additional skew distortions. Pitch, the rotation of the aircraft or spacecraft in the direction of motion, changes the scale across lines nonlinearly and causes aspect distortions. Roll is the rotation about the velocity vector (fig. 2.14*b*). It introduces scale changes in the line direction similar to panoramic distortion. Attitude effects are a major cause of geometric distortion in scanning camera images because of the serial nature of the scanning operation. The possibility of this distortion

may cause serious problems for aircraft scanners with the possibility of sudden and irregular attitude changes. The geometry of frame camera images is internally consistent.

5. *Altitude changes*—These changes cause scale errors.
6. *Perspective errors*—These errors can occur if the image data result from a perspective projection. The effect is similar to a linearly varying scale factor error.

The effects of these geometric errors are shown in figure 2.15. Geometric distortions in remotely sensed images are discussed in [20] and [21]. Techniques to correct for geometric distortions are discussed in section 3.3.

2.4.2. Radiometric Point Degradation

In some imaging systems the object brightness is not uniformly mapped into the image plane. For example, light that passes along the axis of an optical system is generally attenuated less than light that passes through the system obliquely. This degradation is called vignetting. The photocathode of a vidicon is not equally sensitive at all locations, resulting in an image in which equal gray levels do not correspond to equally bright points in the scene. This degradation is known as shading. Such point degradations without spatial blurring may be represented by the following PSF:

$$h(x-\xi, y-\eta) = e(x, y)\delta(x-\xi, y-\eta) \tag{2.88}$$

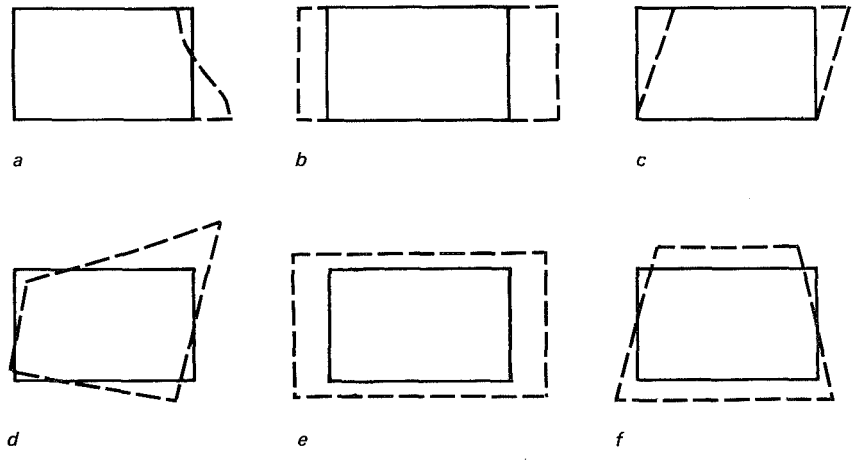


FIGURE 2.15—*Geometric distortions. Solid figures are the correct images, and dashed figures are the distorted images. (a) Scan nonlinearity. (b) Panoramic and roll distortions. (c) Skew (from rotation of Earth). (d) Rotation and aspect distortions (attitude effects). (e) Scale distortion (altitude effect). (f) Perspective distortion.*

In the absence of geometric distortions and noise, the image formation equation reduces to

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(\xi, \eta) e(x, y) \delta(x-\xi, y-\eta) d\xi\, d\eta = e(x, y) f(x, y) \tag{2.89}$$

Such multiplicative point degradations can be corrected by some of the contrast-modification techniques discussed in chapter 4.

2.4.3 Radiometric Spatial Degradation

If the PSF is a function of the object coordinates ξ and η; i.e., $h=h(x-\xi, y-\eta)$, blurring or loss of resolution occurs because of the integrating effect of this imaging system. If no geometric distortions and noise are present, the image formation equation becomes

$$g(x, y) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} h(x-\xi, y-\eta) f(\xi, \eta) d\xi\, d\eta \tag{2.90}$$

Radiometric spatial degradations are caused by defocused systems, diffraction effects and aberrations in optical systems, atmospheric turbulence, and relative motion between imaging system and object.

For an extremely defocused lens the assumption can be made that the PSF is constant over the shape of the aperture and zero elsewhere. Thus, a defocused lens with circular aperture of radius a has the following PSF in polar coordinates [9]:

$$h(r) = \begin{cases} 1 & r \leq a \\ 0 & r > a \end{cases} \tag{2.91}$$

The optical transfer function is

$$H(\omega) = 2\pi a \frac{J_1(a\omega)}{a\omega} \tag{2.92}$$

where $r=(x^2+y^2)^{1/2}$, $\omega=(u^2+v^2)^{1/2}$, and $J_1(\omega)$ is the first-order Bessel function of the first kind. For a square aperture

$$h(x, y) = \begin{cases} 1 & |x| \leq a, \quad |y| \leq a \\ 0 & \text{elsewhere} \end{cases} \tag{2.93}$$

the OTF is

$$H(u, v) = a^2 \frac{\sin 2\pi a u}{2\pi a u} \frac{\sin 2\pi a v}{2\pi a v} \tag{2.94}$$

A more accurate derivation has to consider the effect of diffraction. The functions $H(\omega)$ and $H(u, v)$ are real, because $h(r)$ and $h(x, y)$ are even functions. Because H may be positive or negative, the phase function $\Phi(u, v)$ of the blurred image will have two values, 0 or π, depending on

the sign of H. The locations $H=0$ where the phase changes are called phase or contrast reversals.

The blur caused by atmospheric turbulence for long exposure times can be approximated by the following PSF [22]:

$$h(x, y) = e^{-b(x^2+y^2)} \tag{2.95}$$

The corresponding optical transfer function is

$$H(u, v) = \frac{1}{b} e^{-(\pi^2/b)(u^2+v^2)} \tag{2.96}$$

Because H is positive, there are no phase reversals in the blurred image.

The degradation caused by a uniform relative motion in the x direction of the form

$$x(t) = V_x t \tag{2.97}$$

between camera and the scene while the image is recorded can be described by

$$h(x, y) = \int_{-T/2}^{T/2} \delta(x - V_x t)\, dx\, dy \tag{2.98}$$

where T is the recording time, and the scene is assumed to be invariant in time. The OTF is given by [4]:

$$H(u, v) = \frac{\sin \pi V_x T u}{\pi V_x u} \tag{2.99}$$

Alternate lobes of $H(u, v)$ have negative values and a phase reversal of π radians. Methods to correct for spatial degradations are discussed in section 3.4.

The PSF and modulation-transfer function for the degradation given in equation (2.95) with $b=0.008$ are shown in figure 2.16. An example for the imaging process with equation (2.80) is given in figure 2.17. Figure 2.17*a* shows the undegraded image, and figure 2.17*b* shows the degradation caused by the blurring effect of the PSF in figure 2.16 and by additive noise generated from uniformly distributed random numbers.

2.4.4 Spectral and Temporal Differences

Spectral and temporal differences in multi-images can either be degradations, or they can convey new information. Compensation and correction of degradations that are due to misalinements and illumination changes in the component images are the topics of chapters 4 and 5.

2.5 Digitization

After images have been formed and recorded, they must be digitized for processing by a digital computer. Image formation and recording have

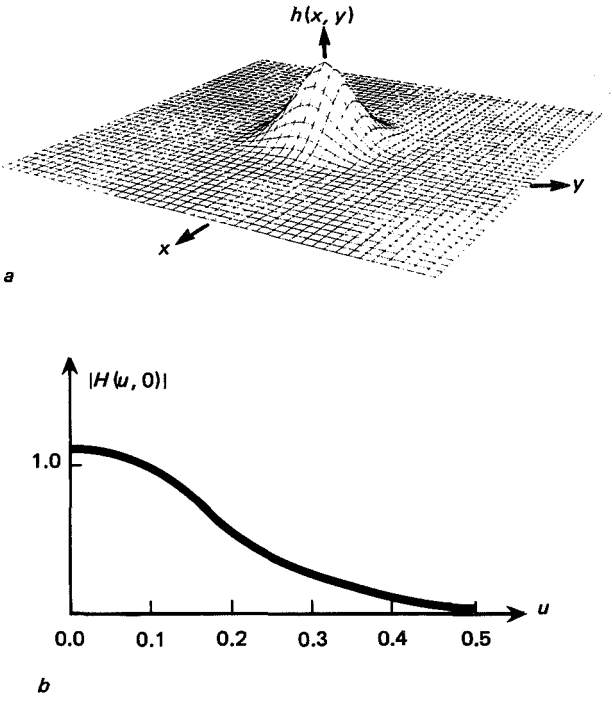


FIGURE 2.16—*Imaging system characteristics. (a) Point-spread function. (b) Modulation transfer function.*

been described in terms of two-dimensional continuous functions $f(x, y)$ and $g(x, y)$, respectively. Digitization consists of sampling the gray level in an image at an M by N matrix of points, and of quantizing the continuous gray levels at the sampled points into K usually uniform intervals. The finer the sampling (M, N large) and the quantization (K large), the better the approximation of the original image.

The aim of sampling and quantization is to represent a continuous image by an array of numbers, called samples, such that a continuous image can be reconstructed from the samples. A digital multi-image with P components is represented by PMN samples. The raster-scan operation of image digitizers and scanners imposes a sequential row structure on the sampled image data. Therefore, the fundamental unit of the data structure is one row of the image matrix. For multi-images the rows of different components may be stored as records in separate files, resulting in a band-sequential (BSQ) storage format. Alternatively, corresponding rows from the P components may be concatenated and stored in one file. This storage format is referred to as band-interleaved by line (BIL). Finally, the values from all components for a given raster point may be

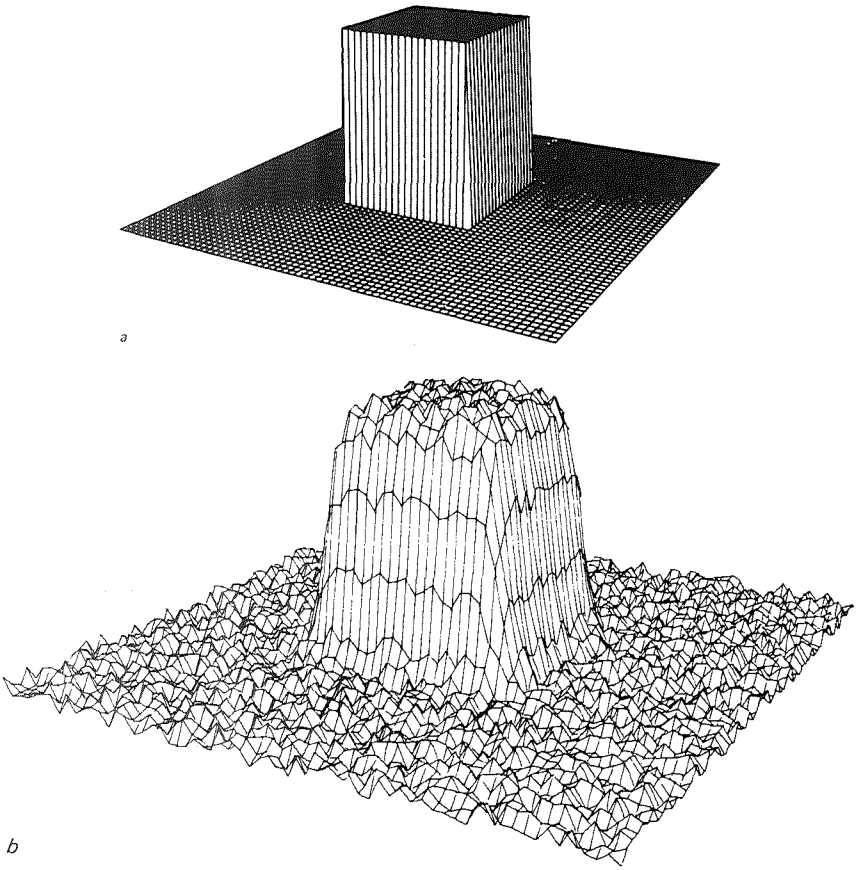


FIGURE 2.17—*Example of image formation. (a) Original scene. (b) Recorded image.*

combined to a P-dimensional vector, and the vectors for one row are concatenated, resulting in one record of the digital image file. This storage format is known as band-interleaved by pixel (BIP).

2.5.1 Sampling

Images are usually sampled at fixed increments $x=j\,\Delta x$, $y=k\,\Delta y$ ($j=1, \ldots, M$, $k=1, \ldots, N$), where Δx and Δy are the sampling intervals in the x and y directions, respectively. The matrix of samples $g(j\Delta x, k\Delta y)$ is the sampled or digital image. In a perfect sampling system the uniform sampling grid is represented by an array of Dirac delta functions [9, 23]:

$$s(x, y)=\sum_{j=-\infty}^{\infty}\sum_{k=-\infty}^{\infty}\delta(x-j\Delta x, y-k\Delta y) \qquad (2.100)$$

The Fourier transform of $s(x, y)$ is $S(u, v)$ [24], where

$$S(u, v) = \frac{1}{\Delta x \Delta y} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} \delta\left(u - \frac{m}{\Delta x}, v - \frac{n}{\Delta y}\right) \tag{2.101}$$

A sampled version g_s of an image $g(x, y)$ is obtained by multiplying $g(x, y)$ with the sampling function $s(x, y)$

$$g_s = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} g(x, y)\delta(x - j\Delta x, y - k\Delta y) \tag{2.102}$$

where g is evaluated at discrete coordinates $(j\Delta x, k\Delta y)$ to form the sampled image g_s.

With the convolution theorem, equation (2.46b), the Fourier transform of the sampled image is

$$G_s(u, v) = \frac{1}{\Delta x \Delta y} \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} G\left(u - \frac{m}{\Delta x}, v - \frac{n}{\Delta y}\right) \tag{2.103}$$

The Fourier transform of the sampled image is a periodic replication of the transform $G(u, v)$. (See fig. 2.18.) This replication of the basic transform introduced by sampling is called aliasing [25]. Aliasing can cause a distorted transform if the replicas overlap. It will be assumed that $g(x, y)$ is a band-limited function. A function $g(x, y)$ is band-limited if its Fourier transform $G(u, v)$ is zero outside a bounded region $|u| > U$, $|v| > V$ in the spatial frequency domain.

The overlap of the replicated spectra can be avoided if $g(x, y)$ is band-limited, and if the sampling intervals are chosen such that

$$\Delta x \leq \frac{1}{2U} \qquad \Delta y \leq \frac{1}{2V} \tag{2.104}$$

The terms $1/(2\Delta x)$ and $1/(2\Delta y)$ are called the Nyquist, or folding frequencies. In physical terms, the sampling intervals must be equal or smaller than one-half the period of the finest detail within the image.

In practical systems the sampling function is not a Dirac delta function, but an array of impulses of finite width. Thus, the sampling array

$$s(x, y) = \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} h_s(x - j\Delta x, y - k\Delta y) \tag{2.105}$$

is composed of $M \times N$ identical, nonoverlapping impulses $h_s(x, y)$, arranged on a grid of spacing Δx, Δy. The sampling impulses are of finite extent; therefore

$$h_s(x, y) = 0 \tag{2.106}$$

outside a resolution cell. The actual values of the image samples are obtained by a spatial integration of the product $g(x, y)s(x, y)$ over each

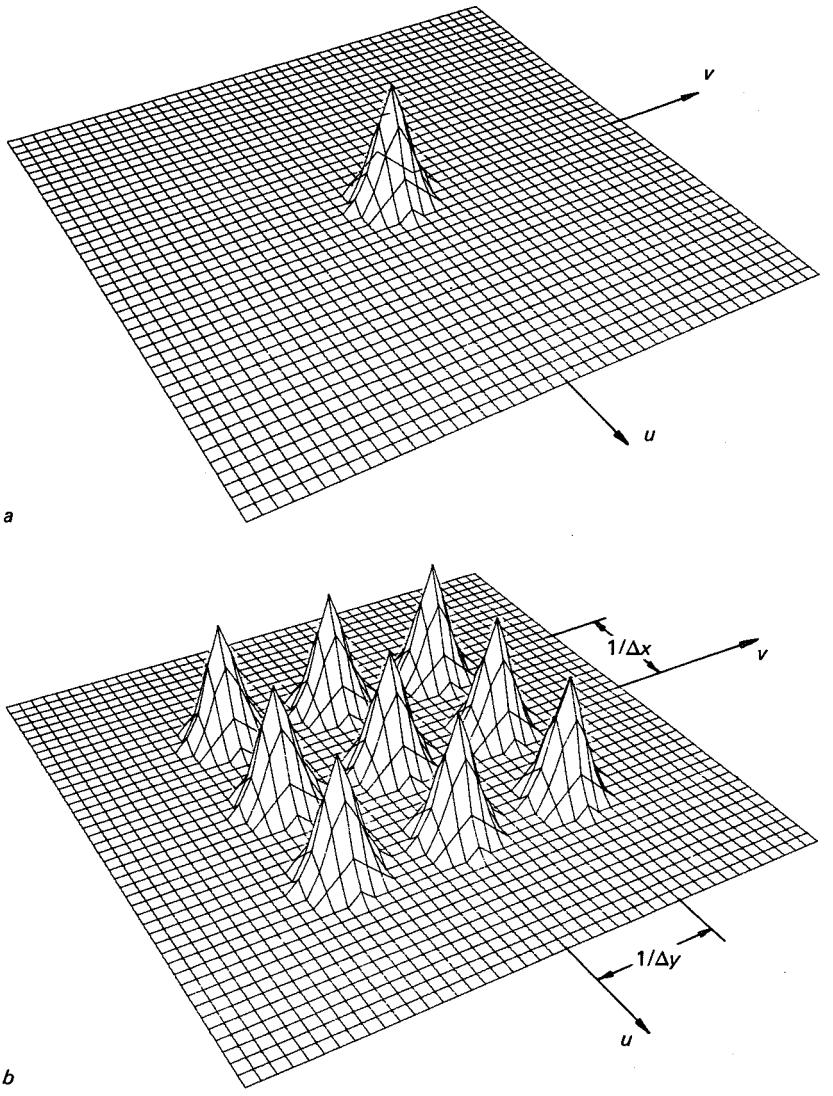


FIGURE 2.18—*Two-dimensional spectrum. (a) Spectrum of band-limited image g. (b) Spectrum of sampled image* g_s^-.

resolution cell [26]. The integration is inherently performed on the image-detector surface. Thus, the sampled image is given by the convolution

$$g_s = \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} g(x, y) h_s(x - j\Delta x, y - k\Delta y) \qquad (2.107)$$

which is evaluated at discrete coordinates $j\Delta x$ and $k\Delta y$. The frequency spectrum G_s of the sampled image now becomes an aliased version of the spectrum, degraded by the convolution with the finite impulse h_s

$$G_s(u, v) = \frac{1}{\Delta x \Delta y} \sum_m \sum_n G\left(u - \frac{m}{\Delta x}, v - \frac{n}{\Delta y}\right) H_s\left(u - \frac{m}{\Delta x}, v - \frac{n}{\Delta y}\right) \tag{2.108}$$

The effects of increasing the sampling intervals and reducing the sampling grid size are shown in figure 2.19. The original image in figure 2.5*a* is an N by N ($N=512$) image with 256 gray levels. Figures 2.19*a* through 2.19*d* show the same image with $N=256$, 128, 64, and 32, respectively.

2.5.2 Quantization

The amplitudes in the sampled image $g_s(j\Delta x, k\Delta y)$ must be divided into discrete values for digital processing. This conversion between analog samples and discrete numbers is called quantization. The number of quantization levels must be sufficiently large to represent fine detail, to avoid false contours in reconstruction, and to match the sensitivity of the human eye. Selective contrast enhancement in digital processing justifies quantization even well beyond the eye's sensitivity.

In most digital image processing systems, a uniform quantization into K_q levels is used. Each quantized picture element is represented by a binary word. If natural binary code is used and the word length is b bits, the number of quantization levels is

$$K_q = 2^b \tag{2.109}$$

The word length b is determined by the signal-to-noise ratio and is chosen to be 6, 7, or 8 bits, resulting in 64, 128, or 256 quantization levels, respectively. The degradation resulting when an image is quantized with an insufficient number of bits is known as contouring effect. This effect, the formation of discrete, rather than gradual, brightness changes, becomes perceptible when $b<6$ [27]. This condition can be improved by nonlinear quantization, which increases the size of quantization intervals that are unlikely to be occupied, and reduces the size of those intervals whose use is highly probable [28]. Nonuniform quantization is also justified on the basis of the properties of the human visual system. In regions with slowly changing gray levels, it is important to have fine quantization. MSS 4, 5, and 6 digital images from Landsats 1 and 2 are logarithmically quantized to 6-bit words onboard the satellite to meet transmission constraints and decompressed to 7-bit words on the ground.

Representing the sampled and quantized picture elements by binary code words is called pulse code modulation (PCM) [29]. A multi-image with $M \times N$ samples, P components, and $K_q = 2^b$ quantization levels requires $PMNb$ bits for its representation using PCM.

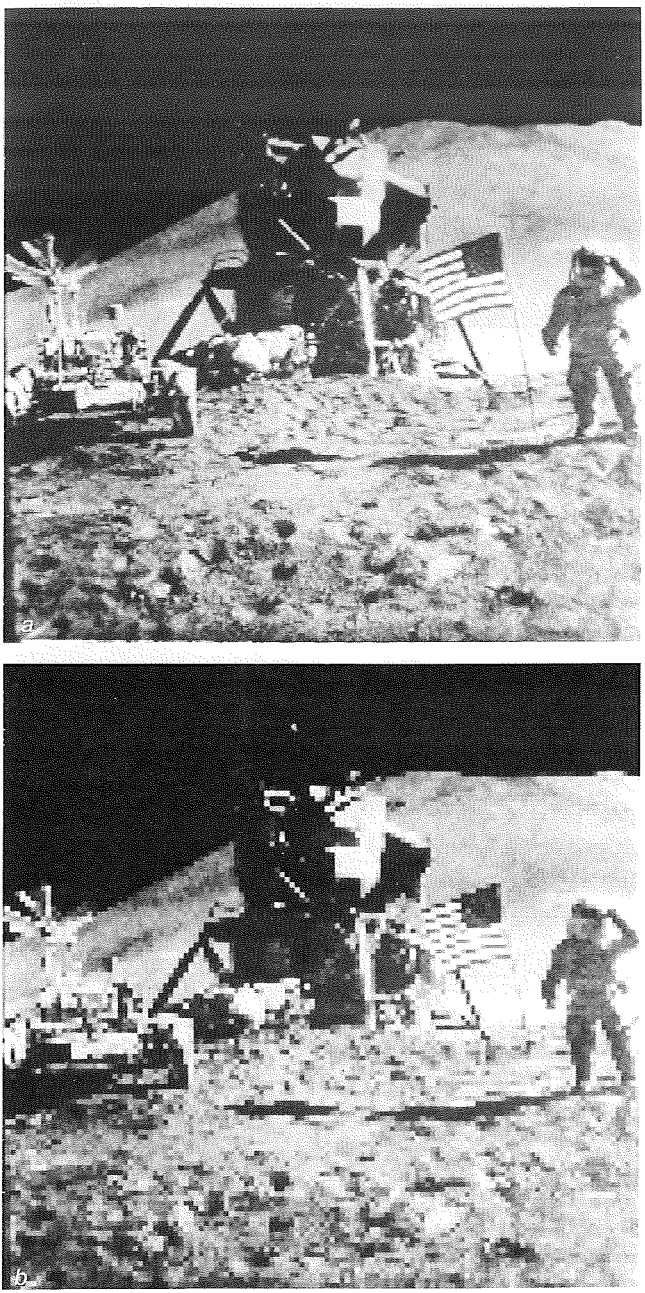


FIGURE 2.19—*Effects of reducing sampling grid size. (a) N = 256. (b) N = 128.*

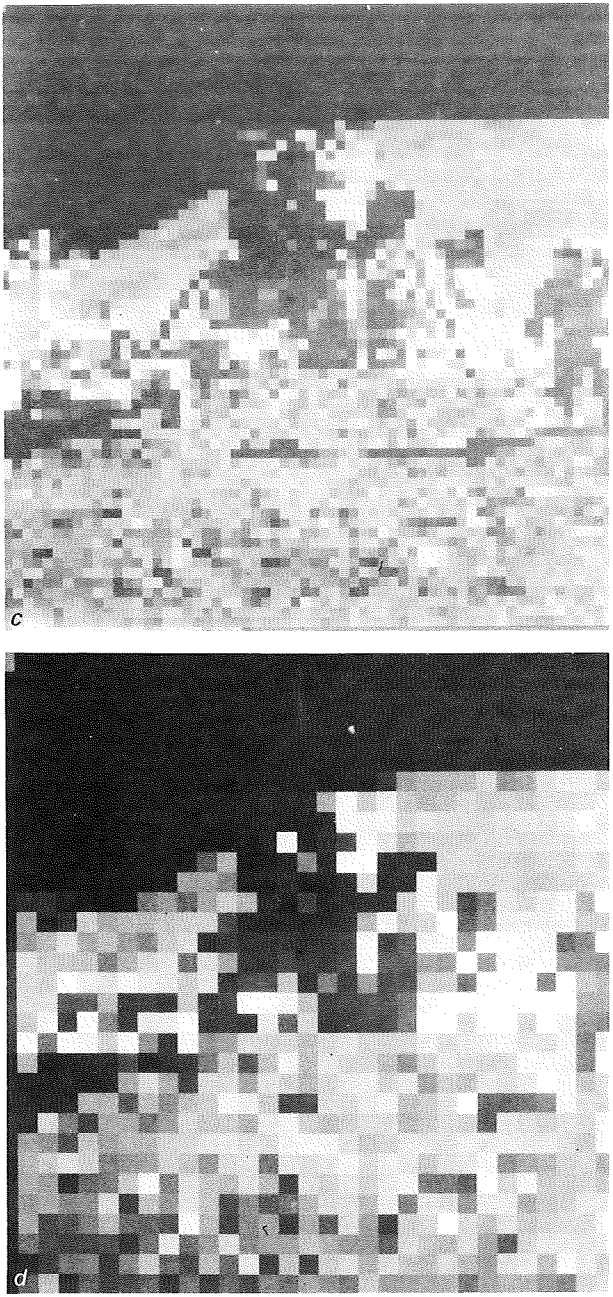


FIGURE 2.19—*Continued. (c) N = 64. (d) N = 32.*

Figure 2.20 illustrates the effects of reducing the number of quantization levels. Figures 2.20*a* through 2.20*d* are obtained by quantizing the image in figure 2.5*a* with $b=6$, 4, 2, and 1 bits while the sampling intervals are kept the same. The false contouring becomes obvious as b is reduced. To obtain a faithful representation of digital images, at least 6 bits are required, and 8 bits are used in general.

2.6 Operations on Digital Images

2.6.1 Discrete Image Transforms

A digital image is represented by an M by N array of numbers, i.e., by a matrix. Matrices will be denoted by uppercase boldfaced letters **A**, **B**, etc., or by $[a]$, $[b]$, etc. Because uppercase letters also denote transforms, the latter notation will primarily be used for matrices. Vectors will be denoted by lowercase boldfaced letters, **a**, **b**, etc.

A discrete representation of the transform pair equations (2.20a) and (2.20b), is given by

$$F(m, n) = \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} \phi_{mn}(j, k) f(j, k)$$
$$m=0, 1, \ldots, M-1 \qquad (2.110a)$$
$$n=0, 1, \ldots, N-1$$

$$f(j, k) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} \phi_{mn}^*(j, k) F(m, n)$$
$$j=0, 1, \ldots, M-1 \qquad (2.110b)$$
$$k=0, 1, \ldots, N-1$$

where $[f]$ and $[F]$ are M by N matrices. $\phi_{mn}(j, k)$ is an element of a four-dimensional operator. Equations (2.110a) and (2.110b) may be written in vector form as

$$\mathbf{F} = \Phi \mathbf{F} \qquad (2.111a)$$

$$\mathbf{f} = \Phi^* \mathbf{f} \qquad (2.111b)$$

The vectors **F** and **f**, each with MN components, are created by lexicographical ordering of the column vectors of matrices $[F]$ and $[f]$, respectively (i.e., the first column of matrix $[f]$ becomes vector elements 1 through M, the second column becomes vector elements $M+1$ through $2M$, . . .). Φ is an MN by MN matrix.

The transformation matrices ϕ_{mn} are said to be separable if for each (m, n),

$$\phi_{mn}(j, k) = p_m(j) q_n(k) \qquad (2.112)$$

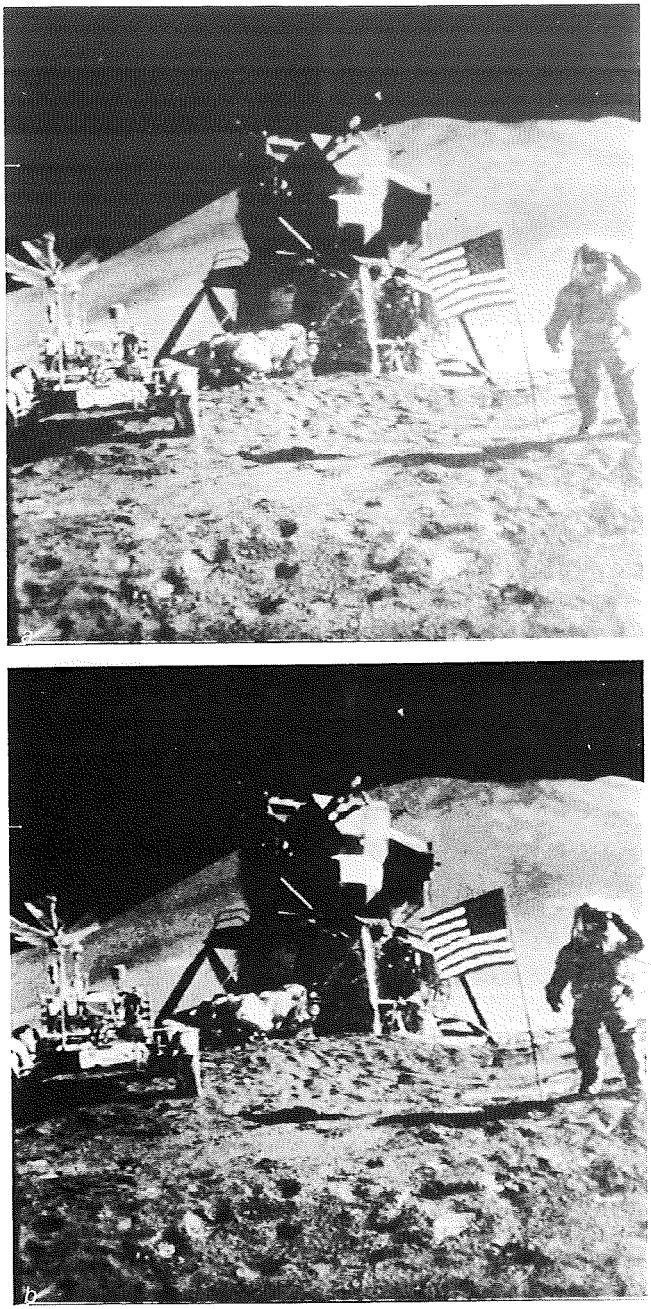


FIGURE 2.20—*Effects of reducing quantization levels. (a) b = 6. (b) b = 4.*

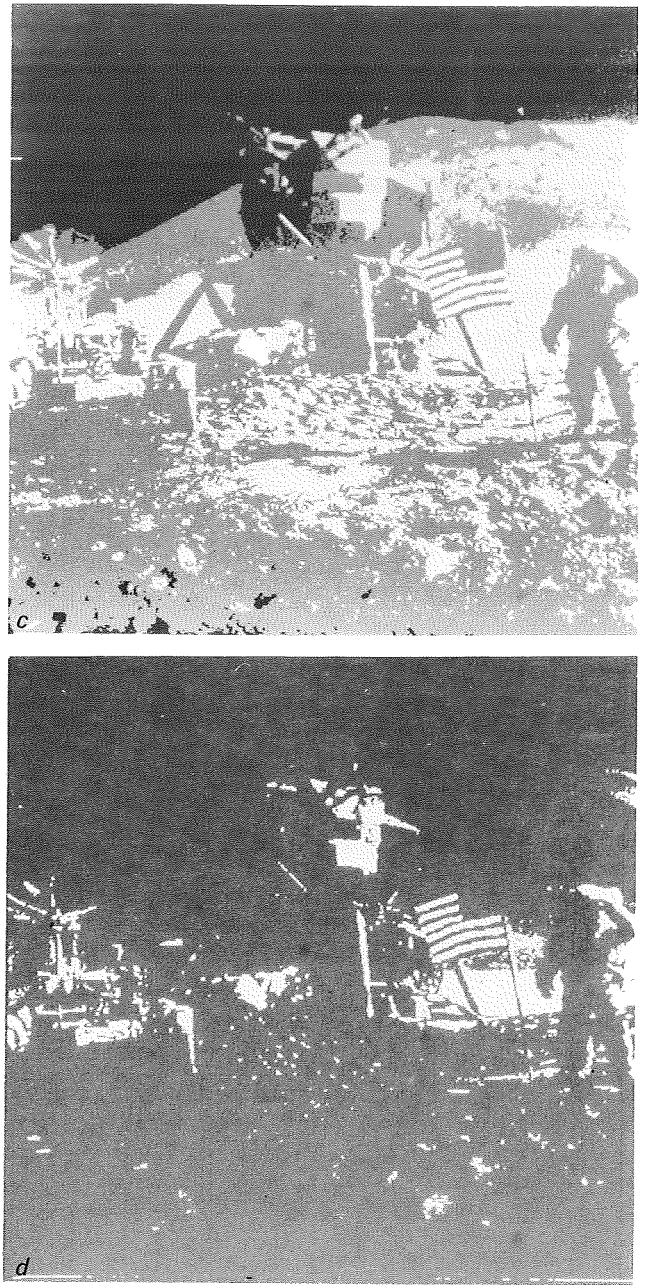


FIGURE 2.20—*Continued. (c) b = 2. (d) b = 1.*

A separable two-dimensional transform can be computed as a sequence of one-dimensional transforms. First, the rows of $f(j, k)$ are transformed to

$$F(m, k) = \sum_{j=0}^{M-1} f(j, k) p_m(j) \tag{2.113}$$

followed by the transformation of the columns of $F(m, k)$ to

$$F(m, n) = \sum_{k=0}^{N-1} F(m, k) q_n(k) \tag{2.114}$$

With the M by M matrix **P**, where

$$P(m, j) = p_m(j)$$

and the N by N matrix **Q**, where

$$Q(n, k) = q_n(k)$$

equation (2.110b) can be written as

$$[F] = \mathbf{P}^T[f]\mathbf{Q} \tag{2.115a}$$

Because **P** and **Q** are unitary matrices, the inverse transform is given by

$$[f] = \mathbf{P}^*[F](\mathbf{Q}^*)^T \tag{2.115b}$$

where $\mathbf{P}^*$ denotes the complex conjugate matrix of **P**. Matrix **F** may be considered as the expansion of an image $[f]$ into a generalized spectrum. Each component of the expansion in the transform domain represents the contribution of that orthogonal matrix to the original image. In this context the concept of frequency may be generalized to orthogonal functions other than sine and cosine waveforms [30].

Separable unitary transforms useful for image processing are the Fourier, cosine, and Hadamard transforms. Fast computational algorithms exist for these transforms. The Karhunen–Loéve transform is a non-separable transform with important image-processing applications.

2.6.1.1 Discrete Fourier Transform

For the discrete Fourier transform (DFT), the elements of the unitary transform matrices are given by

$$\left.\begin{aligned} p^{(m)}(j) &= \frac{1}{M} e^{-2\pi ijm/M} \qquad j, m = 0, 1, \ldots, M-1 \\ \text{and} \\ q^{(n)}(k) &= \frac{1}{N} e^{-2\pi ikn/N} \qquad k, n = 0, 1, \ldots, N-1 \end{aligned}\right\} \tag{2.116}$$

The properties of the continuous Fourier transform listed in section 2.2.3.1 also are valid for the discrete transform.

Let $[f]$ represent an M by N matrix of numbers. The DFT, $\mathbf{F}$, of $\mathbf{f}$ is then defined by

$$\left.\begin{aligned} F(m, n) = \frac{1}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f(j, k) \exp -2\pi i\left(\frac{jm}{M} + \frac{kn}{N}\right) \\ m = 0, 1, \ldots, M-1 \\ n = 0, 1, \ldots, N-1 \end{aligned}\right\} \quad (2.117a)$$

The inverse transform is given by

$$\left.\begin{aligned} f(j, k) = \sum_{m=0}^{M-1} \sum_{n=0}^{N-1} F(m, n) \exp 2\pi i\left(\frac{jm}{M} + \frac{kn}{N}\right) \\ j = 0, 1, \ldots, M-1 \\ k = 0, 1, \ldots, N-1 \end{aligned}\right\} \quad (2.117b)$$

The periodicity properties of the exponential factor imply that

$$\left.\begin{aligned} F(m, -n) &= F(m, N-n) \\ F(-m, n) &= F(M-m, n) \\ F(-m, -n) &= F(M-m, N-n) \\ f(j, -k) &= f(j, N-k) \\ f(-j, k) &= f(M-j, k) \\ f(-j, -k) &= f(M-j, N-k) \end{aligned}\right\} \quad (2.118)$$

Therefore, the extensions of $f(j, k)$ and $F(m, n)$ beyond the original domain as given by $[0 \leq (j \text{ and } m) \leq M-1]$ and $[0 \leq (k \text{ and } n) \leq N-1]$ are periodic repetitions of the matrices. This periodicity has important consequences for the computation of the convolution of two M by N matrices $[f]$ and $[h]$ by multiplying their discrete Fourier transforms F and H. This computation is of great practical value in digital image processing, because the DFT can be efficiently computed by the fast Fourier transform (FFT) algorithm [31, 32, 33].

The FFT algorithm assumes that all the data points in the array to be transformed are kept in main storage simultaneously. The size of practical image arrays, however, requires secondary storage, such as magnetic disk or tape. A sequential row-access structure is imposed on the sampled image matrices by the raster-scan operation of scanning instruments and image digitizers. Each access retrieves one row of the image matrix for processing. This structure makes operations on columns difficult. Because the DFT is separable, the rows of the image matrix are transformed and stored as intermediate results in the first step.

In the second step, the columns of the intermediate matrix $F(m, k)$, given in equation (2.113), have to be transformed. This procedure

would result in excessive input and output time, because all rows would have to be read for each column. One way to avoid such an expenditure of effort is to transpose the intermediate matrix $F(m, k)$. An efficient method for matrix transposition when only a part can be kept in main storage and operated on at the same time is described in [34]. A further improvement is achieved by processing several rows at a time and dividing the transposition algorithm into two parts executed when storing and reading blocks of rows of the intermediate transform matrix [35].

In general, the DFT is used to approximate the continuous Fourier transform. It is very important to understand the relationship between the DFT and the continuous transform. The approximation of the continuous transform by the DFT is effected by sampling and truncation. Consider the one-dimensional continuous function $f(x)$ (e.g., a line of an infinite picture $f(x, y)$) and its Fourier transform in figure 2.21*a*. It is assumed that $f(x)$ is band-limited by U. For digital processing $f(x)$ has to be digitized, which is accomplished by multiplication of $f(x)$ with the sampling function $s(x) = \sum \delta(x - j \cdot \Delta x)$. The sampling interval is Δx (see fig. 2.21*b*). The sampled function $f(j\Delta x)$ and its Fourier transform are shown in figure 2.21*c*. This modification of the continuous transform pair caused by sampling is called aliasing [25, 31]. If $\Delta x \leq 1/2U$ there is no distortion of the transform due to aliasing.

Digital processing also requires truncation to a finite number of points. This operation may be represented by multiplication with a rectangular window function $w(x)$, shown in figure 2.21*d*. Truncation causes convolution of the transform $F(u) * S(u)$ with $W(u)$, where $W(u) = (\sin u)/u$, which results in additional frequency components in the transform. This effect is called leakage. It is caused by the side lobes of $(\sin u)/u$ (fig. 2.21*e*). The transform is also digitized with a sampling interval $\Delta u = (M\Delta x)^{-1}$, resulting in $F(m\Delta u)$, which corresponds to a periodic spatial function $f(j\Delta x)$ (fig. 2.21*f*). The discrete transform $F(m\Delta u)$ differs from the continuous transform $F(u)$ by the errors introduced in sampling (aliasing) and spatial truncation (leakage). Aliasing can be reduced by decreasing the sampling interval Δx (if $f(x)$ is not band-limited). Leakage can be reduced by using a truncation function with smaller side lobes in the frequency domain than the rectangular window. A number of different data windows have been proposed for this apodization process. (See sec. 2.2.5.)

The DFT computes a transform $F(m\Delta u)$, $m = 0, 1, \ldots, M-1$, in which the negative half $F(-u)$ is produced to the right of the positive half. This result may be confusing because analysts are accustomed to viewing the continuous transform $F(u)$ from $-U$ to U. For the two-dimensional DFT, the locations of the spatial frequency components (u, v) are shown in figure 2.22. A normal display may be obtained by rearranging the quadrants of the transform matrix as shown in section 2.6.3.

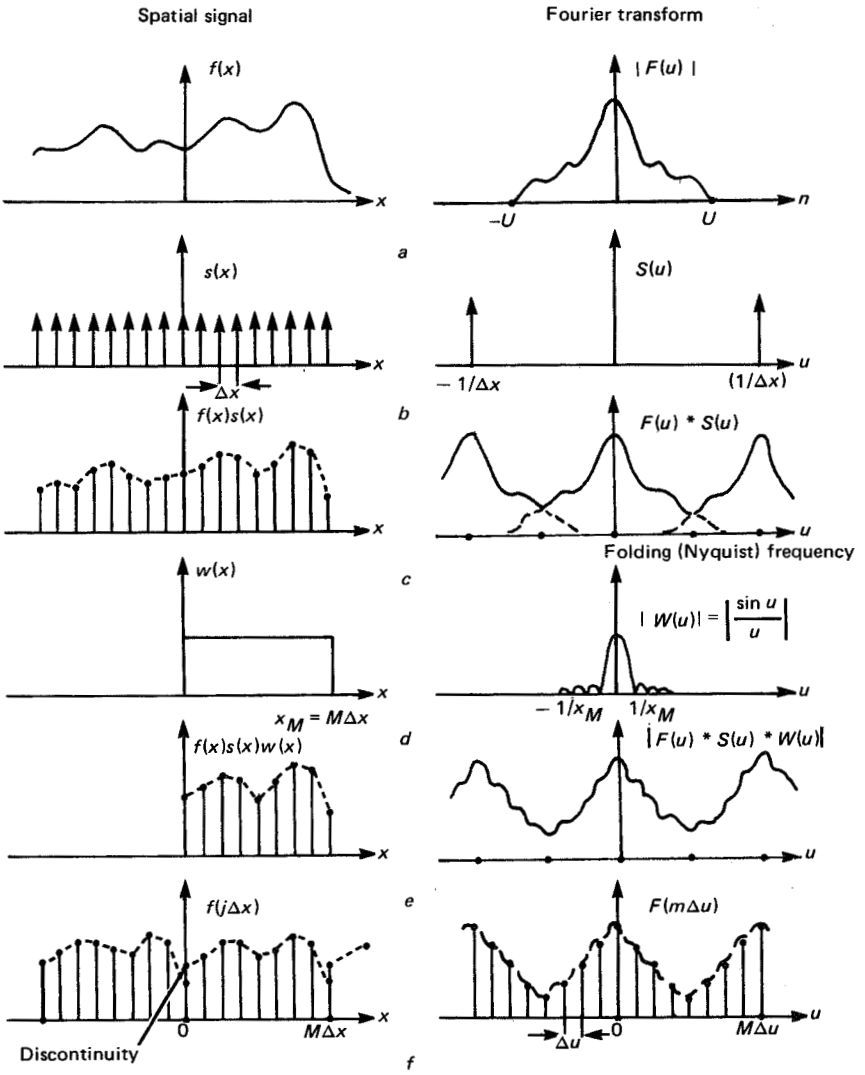


FIGURE 2.21—*Relationship between continuous and discrete Fourier transforms (after Brigham [31]). (a) Continuous transform pair. (b) Sampling function. (c) Sampled. (d) Window function. (e) Sampled and truncated signal. (f) Discrete-transform pair.*

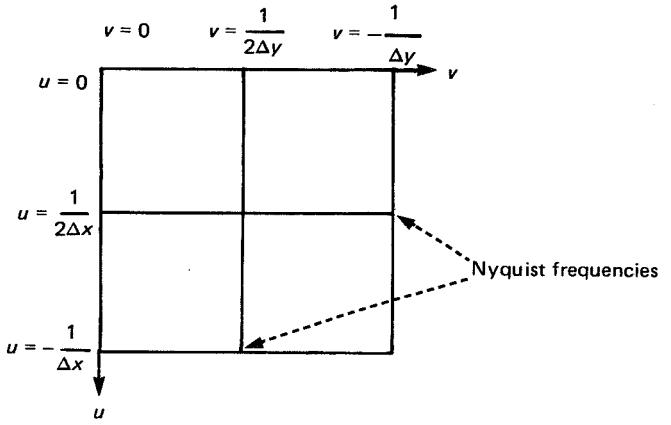


FIGURE 2.22—*Location of spatial frequency components in two-dimensional discrete Fourier transform.*

2.6.1.2 Discrete Cosine Transform

The one-dimensional discrete cosine transform (DCT) of a sequence $f(j)$, $j=1, \ldots, M$ is defined by [36]:

$$\left.\begin{aligned} F(0) &= \frac{\sqrt{2}}{M} \sum_{j=0}^{M-1} f(j) \\ F(m) &= \frac{2}{M} \sum_{j=0}^{M-1} f(j) \cos \frac{2j+1}{2M} \pi m \\ & \qquad m=1, 2, \ldots, M-1 \end{aligned}\right\} \tag{2.119a}$$

The inverse DCT is defined as

$$\left.\begin{aligned} f(j) &= \frac{1}{\sqrt{2}} F(0) + \sum_{m=1}^{M-1} F(m) \cos \frac{2j+1}{2M} \pi m \\ & \qquad j=0, 1, \ldots, M-1 \end{aligned}\right\} \tag{2.119b}$$

The two-dimensional DCT of an image $f(i, j)$ is defined as

$$\left.\begin{aligned} F(0, 0) &= \frac{2}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f(j, k) \\ F(m, n) &= \frac{4}{MN} \sum_{j=0}^{M-1} \sum_{k=0}^{N-1} f(j, k) \cos\left(\frac{2j+1}{2M} \pi m\right) \cos\left(\frac{2k+1}{2N} \pi n\right) \\ & \qquad m=1, 2, \ldots, M-1 \\ & \qquad n=1, 2, \ldots, N-1 \end{aligned}\right\} \tag{2.120a}$$

The two-dimensional inverse DCT is defined as

$$f(j,k)=\frac{1}{2}F(0,0)+\sum_{m=0}^{M-1}\sum_{n=0}^{N-1}F(m,n)\cos\left(\frac{2j+1}{2M}\pi m\right)\cos\left(\frac{2k+1}{2N}\pi n\right)\Bigg\} \quad \begin{matrix} j=0,1,\ldots,M-1 \\ k=0,1,\ldots,N-1 \end{matrix} \tag{2.120b}$$

The two-dimensional DCT is separable and can, therefore, be obtained by successive one-dimensional transformations. In addition, the DCT can be computed with the FFT algorithm [36]. The DCT is primarily used for image compression. (See ch. 9.)

2.6.1.3 Hadamard Transform

If the transform matrices **P** and **Q** in equations (2.115a) and (2.115b) are Hadamard matrices, then $[F]$ is called the Hadamard transform of $[f]$. A Hadamard matrix is a symmetric matrix with elements $+1$ and -1 and mutually orthogonal rows and columns. For Hadamard matrices of order $M=2^l$, the two-dimensional Hadamard transform of an M by M image matrix is defined as

$$F(m,n)=\frac{1}{M^2}\sum_{j=0}^{M-1}\sum_{k=0}^{M-1}f(j,k)(-1)^{r(j,k,m,n)} \tag{2.121}$$

where

$$r(j,k,m,n)=\sum_{i=0}^{M-1}(m_i j_i+n_i k_i)$$

The terms m_i, n_i, j_i, and k_i are the binary representations of m, n, j, and k, respectively [30]. In the context of image processing, the Hadamard transform is primarily used for image compression. (See ch. 9.)

2.6.1.4 Discrete Karhunen–Loéve Transform

The Karhunen–Loéve transform is an orthogonal expansion dependent on the statistical image characteristics. Section 2.2 introduced the concept of representing an image f as a sample of a two-dimensional random field or stochastic process **f**. The mean vector and the correlation matrix that statistically describe the images can be computed in the spectral/temporal dimension, or in the spatial dimension. In the first case, the elements of a multi-image are given by P-dimensional vectors **f** whose elements are the pixel values of the spectral/temporal components for a given spatial location (i,j). P is the number of spectral/temporal components. In the second case, one image matrix $f(i,j)$, $i=1,\ldots,M$, $j=1,$

. . ., N of the random field is represented by a vector $\mathbf{f}$ of $P=MN$ dimensions, which is obtained by lexicographically arranging the columns of matrix $[f]$ into vector $\mathbf{f}$. The statistical characteristics of $\mathbf{f}$ (see sec. 2.2.2 for continuous functions) are given by the mean vector

$$\mathbf{m}=E\{\mathbf{f}\} \tag{2.122}$$

and the covariance matrix

$$\mathbf{C}=\mathbf{R}-\mathbf{m}\mathbf{m}^T \tag{2.123}$$

where $\mathbf{R}$ is the correlation matrix

$$\mathbf{R}=E\{\mathbf{f}\mathbf{f}^T\} \tag{2.124}$$

Let $\mathbf{f}$ be a vector in a P-dimensional vector space. Let $\{\mathbf{t}_n\}$ be a complete set of orthonormal vectors in the same space. Then, an arbitrary vector $\mathbf{f}$ may be expanded as

$$\mathbf{f}=\sum_{n=1}^{P} F_n\mathbf{t}_n \tag{2.125}$$

where the coefficients of expansion F_n are given by

$$F_n=\mathbf{t}_n{}^T\mathbf{f} \qquad n=1, \ldots, P \tag{2.126}$$

For some applications (e.g., feature selection for classification or image compression) $\mathbf{f}$ must be approximated in a mean-square-error sense with as few coefficients F_n, $n=1, \ldots, N<P$, as possible. (Here N is different from the number of columns N.)

Therefore, by minimizing the error e, where

$$e=E\left\{ \| \mathbf{f}-\sum_{n=1}^{N} F_n\mathbf{t}_n \|^2 \right\} \tag{2.127}$$

the best basis vectors $\mathbf{t}_n$, $n=1, \ldots, N$ are obtained. $\| \mathbf{f} \|^2=\mathbf{f}^T\mathbf{f}$ is the Euclidean norm of $\mathbf{f}$. With equation (2.125) the error becomes

$$e=E\left\{ \| \sum_{n=N+1}^{P} F_n\mathbf{t}_n \|^2 \right\} \tag{2.128}$$

Using equation (2.126) with the orthonormality property and expanding the norm yield

$$e=\sum_{n=N+1}^{P} \mathbf{t}_n{}^T E\{\mathbf{f}\mathbf{f}^T\}\mathbf{t}_n$$

or

$$e=\sum_{n=N+1}^{P} \mathbf{t}_n{}^T R\mathbf{t}_n \tag{2.129}$$

Because R is a symmetric, positive definitive matrix, equation (2.129) can be minimized with the Lagrange method, which yields

$$R\mathbf{t}_n = \lambda_n \mathbf{t}_n \tag{2.130}$$

Thus, the optimal vectors belonging to $\{\mathbf{t}_n\}$ are the eigenvectors of R, and the values belonging to $\{\lambda_n\}$ are the corresponding eigenvalues. The matrix R is the correlation matrix given in equation (2.124). It has exactly P positive different eigenvalues and P linear independent orthonormal eigenvectors $\mathbf{t}_n$. The minimum error becomes

$$e = \sum_{n=N+1}^{P} \mathbf{t}_n{}^T \lambda_n \mathbf{t}_n = \sum_{n=N+1}^{P} \lambda_n \tag{2.131}$$

where the values λ_n, $n=N+1, \ldots, P$, are the eigenvalues associated with the eigenvectors not included in the expansion equation (2.125). Thus, the approximation error will be minimized if the eigenvectors $\mathbf{t}_n$ corresponding to the N largest eigenvalues are chosen for the representation of $\mathbf{f}$.

The eigenvectors, ordered according to the decreasing magnitude of their corresponding eigenvalues ($\lambda_1 > \lambda_2 > \ldots \lambda_P$), can be combined into the P by P transform matrix $\mathbf{T}$, given by

$$\mathbf{T} = \begin{pmatrix} \mathbf{t}_1{}^T \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{t}_P{}^T \end{pmatrix} \tag{2.132}$$

A pixel vector of the transformed image (the vector of the coefficients of expansion) is then given by

$$\mathbf{F} = \mathbf{T}\,\mathbf{f} \tag{2.133}$$

A reduced transform matrix $\mathbf{T}_n$ may be defined by combining the $N<P$ ordered eigenvectors belonging to the N largest eigenvalues into the N by P matrix

$$\mathbf{T}_N = \begin{pmatrix} \mathbf{t}_1{}^T \\ \cdot \\ \cdot \\ \cdot \\ \mathbf{t}_N{}^T \end{pmatrix} \tag{2.134}$$

An N-dimensional reduced pixel vector (feature vector) is then computed by

$$\mathbf{F}_n = \mathbf{T}_n\,\mathbf{f} \tag{2.135}$$

Because **T** is computed from a sample covariance matrix, the transformation will be different for each application. Equations (2.130) and (2.133) define the discrete Karhunen–Loéve transform or transform to principal components [10].

The correlation matrix of the transformed multi-images (the principal component images) is

$$\mathbf{R}_F = E\{\mathbf{F}\,\mathbf{F}^T\} = \mathbf{T}\,\mathbf{R}\,\mathbf{T}^T = \Lambda \tag{2.136}$$

where $\Lambda = \text{diag}\ (\lambda_n)$ is a diagonal matrix consisting of the ordered eigenvalues λ_n of **R**. Thus, the principal components are uncorrelated, and each eigenvalue λ_n is the variance of the nth principal component:

$$\lambda_n = (\sigma_n{}^F)^2 \tag{2.137}$$

Because **T** is an orthogonal transform, the total data variance is preserved; i.e.,

$$\sum_{n=1}^{P} (\sigma_n{}^f)^2 = \sum_{n=1}^{P} \lambda_n$$

where $(\sigma_n{}^f)^2$ is the variance of the nth original image component. It is evident that the ordering of the eigenvectors in equation (2.132) insures that each principal component has variance less than the previous component. The Karhunen–Loéve transform is usually applied in the spectral/temporal dimension of a multi-image for feature selection, enhancement, and image compression.

The Fourier, cosine, and Hadamard transform are expansions independent of the images and are, therefore, not optimal in the sense of yielding an approximation with a minimum number of uncorrelated coefficients. The Fourier transform yields an optimal expansion only if the images have statistical characteristics such that (1) the diagonal elements of their covariance matrix are all equal, and (2) the correlation between picture elements i and j is only a function of $j-i$. These characteristics are satisfied by a Markov covariance matrix, which in some cases is representative of image properties [7]. The Fourier, cosine, and Hadamard transforms are attractive because of the existence of fast algorithms for their implementation [30]. They are used for image compression in the spatial dimension. The principal application of the Fourier transform is in filtering for image restoration and enhancement.

2.6.2 Discrete Convolution

Linear space invariant image formation and linear filtering are convolution operations. A discrete representation of the convolution integral, equation (2.80), is obtained by approximate integration, where the continuous functions are described by samples, spaced over a uniform grid Δx, Δy.

$$g(j\Delta x, k\Delta y)$$

$$\approx \Delta x\,\Delta y \sum_{m=j}^{j+K-1} \sum_{n=k}^{k+L-1} w_{mn}\, f(m\Delta x, n\Delta y)\, h([j-m+K]\,\Delta x, [k-n+L]\Delta y) + n(j\Delta x, k\Delta y) \tag{2.138}$$

where w_{mn} is one of the integration coefficients, $M=x_m/\Delta x$, $N=y_m/\Delta y$, $K=x_h/\Delta x$, and $L=y_h/\Delta y$. For $\Delta x=\Delta y=1$ and $w_{mn}=1$, equation (2.138) can be written as

$$g(j, k)=\sum_{m=0}^{K-1} \sum_{n=0}^{L-1} f(m, n)\, h(j-m, k-n)+n(j, k) \tag{2.139}$$

$$j=K, K+1, \ldots, M-1$$
$$k=L, L+1, \ldots, N-1$$

In this discrete representation, $[f]$, $[h]$, $[g]$, and $[n]$ are matrices, formed by sampling the corresponding continuous functions, of the following sizes: $[f]$ is of size M by N, $[h]$ is of size K by L, and $[g]$ and $[n]$ are of size M' by N', $(M'=M-K;\ N'=N-L)$. Equation (2.139) is a linear system of equations that can be written in vector form as

$$\mathbf{g}=\mathbf{B}\mathbf{f}+\mathbf{n} \tag{2.140}$$

where $\mathbf{g}$ and $\mathbf{n}$ are vectors with $M'N'$ components each, and $\mathbf{f}$ is a vector with MN components, created by lexicographically ordering the column vectors of matrices $[g]$, $[n]$, and $[f]$, respectively. The matrix $\mathbf{B}$ has dimensions $M'N'$ by MN and can be partitioned as

$$\mathbf{B}=\begin{pmatrix} \mathbf{B}_{1,1} & \mathbf{B}_{1,2} & \ldots & \mathbf{B}_{1,L} & 0 & 0 & \ldots & & 0 \\ 0 & \mathbf{B}_{2,2} & \ldots & \mathbf{B}_{2,L} & \mathbf{B}_{2,L+1} & 0 & \ldots & & \\ & & & & & & & & \\ 0 & 0 & \ldots & & 0 & \mathbf{B}_{N',N-L+1} & \ldots & \mathbf{B}_{N',N} & \end{pmatrix} \tag{2.141}$$

The structure of $\mathbf{B}$ is determined by $\mathbf{B}_{r,s}=\mathbf{B}_{r-1,s-1}$. Each submatrix $\mathbf{B}_{r,s}$ is a circulant matrix. In matrices with this property, each row is equal to the row preceding it shifted one element to the right, with the last element wrapped around to the first place. Circulant matrices have the special property that they are diagonalized by the DFT [37].

The computation of equation (2.139) requires $MNKL$ operations. The convolution theorem, equation (2.46a), permits computation of equation (2.139) in the spatial frequency domain with the FFT algorithm [32, 33]. Whenever the values of $f(j, k)$ and $h(j, k)$ are required for indices outside the ranges $0\leq j\leq M-1$, $0\leq k\leq N-1$ and $0\leq j\leq K-1$, $0\leq k\leq L-1$, respectively, they must be obtained by the rules given in equation (2.118).

With this condition, equation (2.139) becomes a periodic or circular convolution.

To avoid distortion of the convolution due to wraparound, the images are extended with zeroes. Extended matrices $[f_e]$, $[h_e]$, $[g_e]$, and $[n_e]$ of common size P by Q are defined according to

$$\left.\begin{aligned} f_e(j,k) &= \begin{cases} f(j,k) & \text{for } 0\le j\le M-1,\ 0\le k\le N-1 \\ 0 & \text{for } M\le j\le P,\ N\le k\le Q \end{cases} \\ h_e(j,k) &= \begin{cases} h(j,k) & \text{for } 0\le j\le K-1,\ 0\le k\le L-1 \\ 0 & \text{for } K\le j\le P,\ L\le k\le Q \end{cases} \\ g_e(j,k) &= \begin{cases} g(j,k) & \text{for } 0\le j\le M'-1,\ 0\le k\le N'-1 \\ 0 & \text{for } M'\le j\le P,\ N'\le k\le Q \end{cases} \\ n_e(j,k) &= \begin{cases} n(j,k) & \text{for } 0\le j\le M'-1,\ 0\le k\le N'-1 \\ 0 & \text{for } M'\le j\le P,\ N'\le k\le Q \end{cases} \end{aligned}\right\} \tag{2.142}$$

where $P=2^p\ge M+K-1$, $Q=2^q\ge N+L-1$ (p, q integer). If these inequalities are satisfied, $h(j-m, k-n)$ will never wrap around and engage a nonzero portion of $f(m, n)$, and therefore the circular convolution will be identical to the desired linear convolution.

The number of computational operations required to obtain the convolution with the FFT is on the order of PQ ($2\log P+2\log Q+1$). If multiplications by zero are avoided, the convolution in the transform domain can be more efficient than direct computation. Figure 2.23 shows a comparison of computer times for direct and Fourier transform convolutions for a 256 by 256 image for different sizes of the matrix $[h]$ with an IBM 360/75 computer. The figure suggests that indirect computation should be used for filter sizes greater than $K=L=13$.

It is important to note that the matrices $[f_e]$ and $[h_e]$ must be of the same size, P by Q. If the matrix $[h]$ is much smaller than $[f]$ ($K\ll M$, $L\ll N$), block filtering by convolving segments of f with h can be used [38].

2.6.3 Discrete Crosscorrelation

The discrete crosscorrelation between an M by N image $f(j, k)$ and P by Q image $g(j, k)$ is defined by

$$R(r,s)=\sum_{j=1}^{M}\sum_{k=1}^{N} f(j,k)\, g(j+r, k+s) \tag{2.143}$$

$R(r, s)$ may be computed directly or by the DFT with the property given by equation (2.47). Because both $f(j, k)$ and $g(j, k)$ are assumed to be periodic two-dimensional sequences for the indirect computation, they may be extended by zeroes as in equation (2.142) to avoid wraparound and distortion of the correlation function. In many image processing applications f is smaller than g ($M<P$, $N<Q$), and P, Q may be chosen

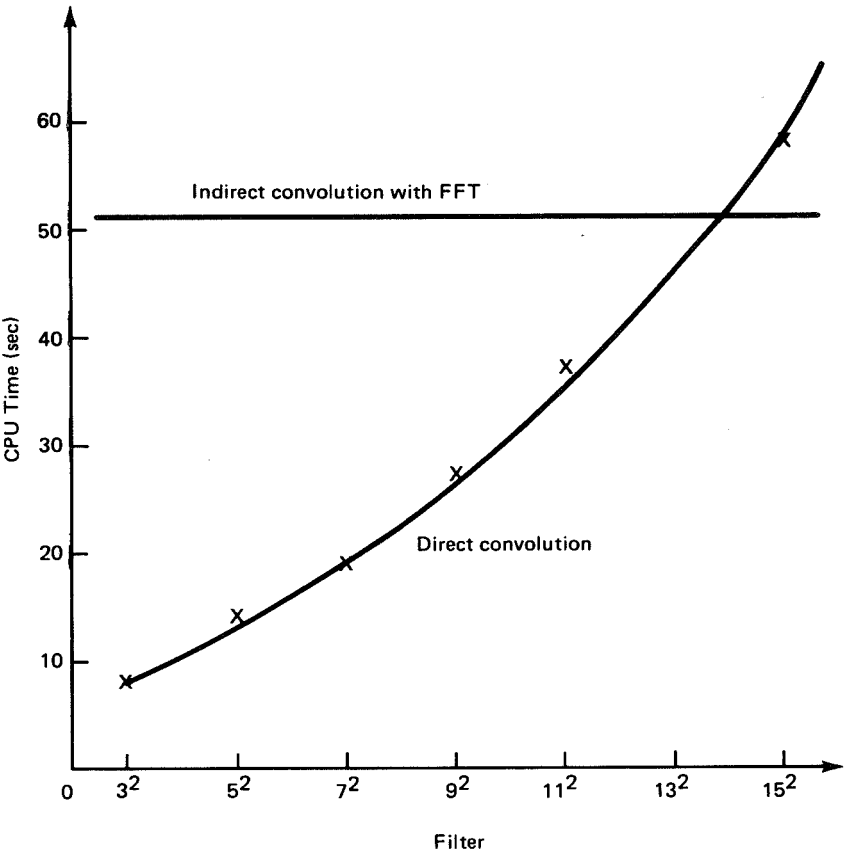


FIGURE 2.23—*Computation time for discrete convolution (IBM 360/75).*

such that $P=2p$, $Q=2q$ (p, q integer). In this case only f will be extended by zeroes to a P by Q matrix:

$$f_e(j,k)=\begin{cases} f(j,k) & 0\leq j\leq M-1 \quad 0\leq k\leq N-1 \\ 0 & M\leq j\leq P-1 \quad N\leq k\leq Q-1 \end{cases} \tag{2.144}$$

The DFTs $F_e(m,n)$ and $G(m,n)$ of $f_e(j,k)$ and $g(j,k)$ are computed with equation (2.117a), and the crosscorrelation function is determined by the inverse transform

$$R(r,s)=\mathcal{F}^{-1}\{F^*(m,n)\,G(m,n)\} \tag{2.145}$$

where $\mathcal{F}^{-1}$ is computed by equation 2.117b. There is some wraparound in $R(r,s)$, but a valid correlation function of size $(P-M+1)$ by $(Q-N+1)$ is contained in $R(r,s)$. The correlation value for no shift is the point $m=0$, $n=0$, and positive shifts in both directions are given

for $r=1$, $s=1$ up to $(P-M)/2$, $(Q-N)/2$. The point $(P+M)/2$, $(Q+N)/2$ represents maximum negative shift in both directions, and $R(P-1, Q-1)$ is the correlation value for a negative one-element shift in both directions. (See fig. 2.24.) The remaining values for $r=[(P-M)/2]+1$ to $[(P+M)/2]-1$ and $[(Q-N)/2]+1$ to $[(Q+N)/2]-1$ are invalid. If $R(r, s)$ is plotted as a correlation surface, zero displacement appears in the upper left corner. A more conventional display has zero shift at the center, with lines and samples starting at 1. Thus, the valid quadrants of R have to be extracted and rearranged to yield a correlation function with zero shift in the center. The transformation

$$R(r', s')=R(r, s)$$

where

$$\left.\begin{aligned} r'&=\begin{cases}\dfrac{P-M}{2}-r+1 & r=0, 1, \ldots, \dfrac{P-M}{2}\\[2ex] \dfrac{3P-M}{2}-r+1 & r=\dfrac{P+M}{2}, \ldots, P-1\end{cases}\\ s'&=\begin{cases}s+\dfrac{Q-N}{2}+1 & s=0, 1, \ldots, \dfrac{Q-N}{2}\\[2ex] s-\dfrac{Q+N}{2}+1 & s=\dfrac{Q+N}{2}, \ldots, Q-1\end{cases}\end{aligned}\right\} \quad (2.146)$$

yields a $(P-M+1)$ by $(Q-N+1)$ correlation function with zero shift (no displacement of f and g) at $[(P-M)/2]+1$, $[(Q-N)/2]-1$.

2.7 Reconstruction and Display

Image display is concerned with the regeneration of continuous pictures from sampled images. Display systems use CRTs or write directly on film.

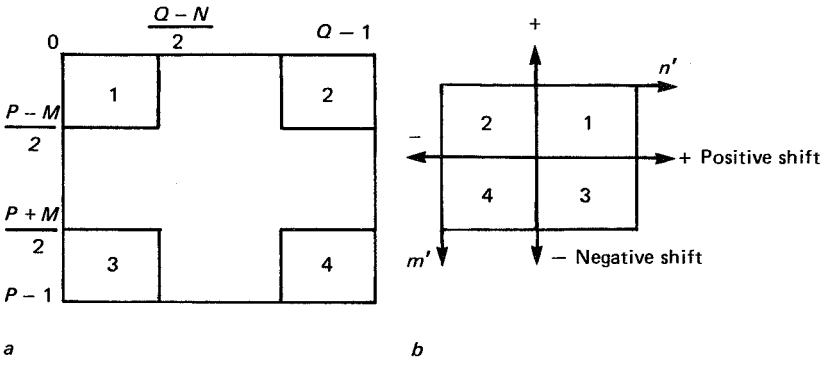


FIGURE 2.24—*Two-dimensional crosscorrelation computed by discrete Fourier transform. (a) Computed correlation matrix. (b) Rearranged, valid correlation matrix.*

A light spot of finite size is focused and projected by optics onto the film or CRT surface. The spot intensity is modulated, and the spot sweeps across the display plane in a raster-scan fashion to create a continuous picture.

A continuous picture may be obtained from the samples by spatial interpolation or filtering. Let $h_d(x, y)$ denote the impulse response of the interpolation filter, and $H_d(u, v)$, its transfer function. The reconstructed continuous picture $g_d(x, y)$ is obtained by a convolution of the digital image $g_s(i, j)$ with the reconstruction filter or display spot impulse response h_d:

$$g_d(x, y) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} g_s(j\Delta\bar{x}, k\Delta\bar{y})\, h_d(x - j\Delta\bar{x}, \bar{y} - k\Delta\bar{y}) \quad (2.147)$$

where $\Delta x = a\Delta x$ and $\Delta y = b\Delta y$ are the display spot spacings. The frequency spectrum of the reconstructed and displayed image is (see eq. (2.108) and (2.42) and ref. [26]):

$$G_d(u, v) = \frac{1}{a\,b} G_s\left(\frac{u}{a}, \frac{v}{b}\right) H_d(u, v)$$

$$= H_d(u, v) \frac{1}{\Delta\bar{x}\,\Delta\bar{y}} \sum_m \sum_n G\left(\frac{u - \frac{m}{\Delta x}}{a}, \frac{v - \frac{n}{\Delta y}}{b}\right) H_s\left(\frac{u - \frac{m}{\Delta x}}{a}, \frac{v - \frac{n}{\Delta y}}{b}\right) \quad (2.148)$$

assuming that the spectrum G_s was not modified by digital processing. The difference between the sampling and display spot spacings is reflected in the scaled spectrum G_d, where equation (2.42) is used.

Equation (2.148) shows the aliasing and the degradations caused by sampling and display. It is evident that the spectrum of the reconstructed image could be made equal to the spectrum of the original image g, if no aliasing were present, if sampling would not degrade the spectrum, and if the reconstruction filter H_d would select the principal Fourier transform $G(u, v)$ with $m = n = 0$, and reject all other replications in the frequency domain.

The first condition is met by a band-limited image if the sampling intervals are chosen according to equation (2.104). The second condition is met if the sampling impulse is an ideal delta function. The third condition is met if the reconstruction filter transfer function is

$$H_d(u, v) = \begin{cases} 1 & \text{for } |u| \leq U \quad \text{and} \quad |v| \leq V \\ 0 & \text{elsewhere} \end{cases} \quad (2.149)$$

The impulse response of this reconstruction filter (ideal low-pass filter) is

$$h_d(x, y) = U\,V \frac{\sin 2\pi Ux}{2\pi Ux} \frac{\sin 2\pi Vy}{2\pi Vy} \quad (2.150)$$

Thus, the conditions for exact image reconstruction are that the original image is band-limited, that it is spatially sampled at a rate twice its highest spatial frequency, that the sampling impulse is a delta function, and that the reconstruction filter is designed to pass the frequency spectrum at $m=n=0$ without distortion, and reject all other spectra for which $m, n \neq 0$.

Practically, images are not band-limited because they contain edges and noise, which cause high spatial frequency components in the transform. However, the assumption is a reasonable approximation because most of the image energy is contained in the low-frequency region of the spectrum. The aliasing effects can be reduced by the filtering inherent in the sampling process. In practical systems, the sampling impulse is never a Dirac delta function. Consequently, the spectrum of the sampled image is degraded by the transfer function of the sampling spot. The reconstruction function of the display system cannot be a true $(\sin x)/x$ function, because of finite spot size and positive light. Therefore, there is always aliasing, because the display spot does not completely attenuate the replications of the sampled spectrum of equation (2.108). The aliasing effects in display systems consist of Moiré patterns and edge effects. These effects, however, are negligible if 90 to 95 percent of the image energy lies in a region of frequencies below the Nyquist frequency. This criterion is satisfied by most remote sensing images. Moiré patterns, for example, are usually only visible if there are periodic structures with frequencies near the Nyquist limit in the image.

The quality of the displayed image is also influenced by the transfer characteristics of the display system. Pixel values in digital images represent a particular intensity or optical density. The aim is to generate a display image g_d with the same measurable intensity or density as represented by the corresponding digital image g. Therefore, the generally nonlinear transfer characteristic $g_d = d(g)$ of actual display systems must be measured. Before displaying any image, it is first transformed by the inverse characteristic d^{-1} to compensate for the effect of the display system [39]. The inverse transformation is a point operation; it can be implemented in a lookup table in the display system, or in an image processing function.

2.8 Visual Perception

The designer and user of image processing algorithms and displays has to consider the characteristics of the human visual system. The application of mathematical–statistical techniques to image processing problems frequently requires a measure of image fidelity and quality. For example, in radiometric restoration and image compression a criterion is needed to measure the closeness of the reconstructed image with the original. For visual interpretation it is important to know how the human eye sees an image to determine the best enhancement and display parameters. The

distortion due to the limited dynamic range of display devices can be avoided by consideration of the properties of the visual system [40, 41].

Image analysis is based on the assumption that the information of significance to the human observer may be characterized in terms of the properties of perceived objects or patterns [42]. These properties are determined by the observable psychophysical parameters contrast, contour, texture, shape, and color. Although machine analysis of images relies entirely on features measurable in a picture, visual interpretation consists of a sequence of visual processing stages from the initial detection of objects to the final recognition. Objects or patterns may only be detected when they can be distinguished from their surroundings with respect to the psychophysical parameters. Most of the information about objects resides within their border. Thus, objects may be detected by differing contrast, texture, or color. Recognition depends to a degree on information not resident in or derivable from an image, namely information based on prior experience of the analyst.

The properties of the human visual system determine the useful range and distribution of the psychophysical parameters. On the other hand, the objectives of the image analysis task determine which features in the image are to be represented by the parameters. The detection of objects is aided by image processing. Enhancement techniques are used to increase the contrast, and to enhance contours, texture, and colors.

2.8.1 Contrast and Contour

Contrast is a local difference in luminance and may be defined as the ratio of the average gray value of an object to the average gray level of its background. The sensitivity of the human visual system depends logarithmically on light intensities that enter the eye. Thus, the greater the brightness, the greater the contrast between objects must be to detect any differences. This relationship is known as the Weber–Fechner law.

The apparent brightness of objects depends strongly on the local background intensity. This phenomenon is called simultaneous contrast. In figure 2.25, the small squares have equal intensity, but they appear to have different brightnesses, because their backgrounds have widely differing intensities.

The ability of the visual system to detect sharp edges that define contours of objects is known as acuity. The eye possesses a lower sensitivity for slowly and rapidly varying patterns, but the resolution of midrange spatial frequencies is excellent. Thus, the visual system behaves like a band-pass filter in its ability to detect fine spatial detail. This characteristic is demonstrated in figure 2.26, where the spatial frequency of the sinusoidal pattern increases to the right while the contrast increases downward. The curve along which the pattern is just visible represents the

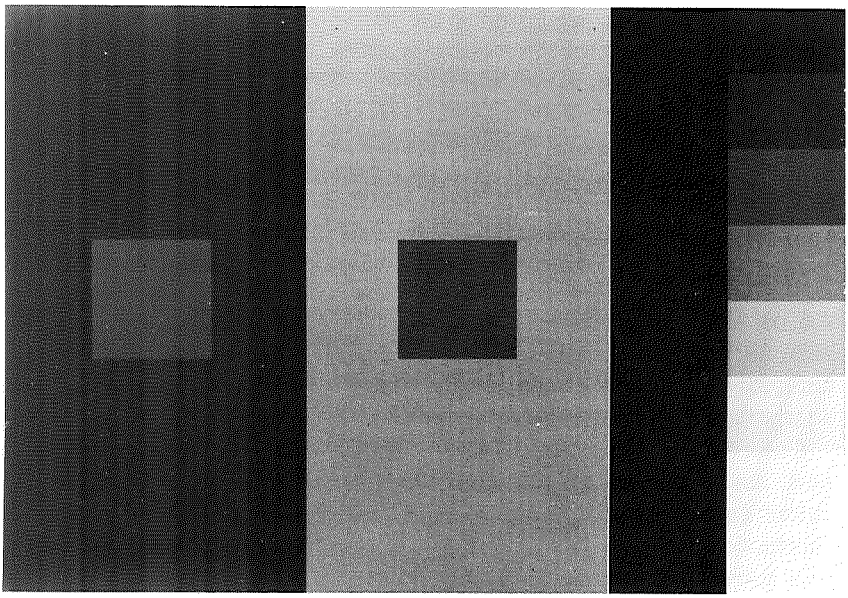


FIGURE 2.25—*Simultaneous contrast.*

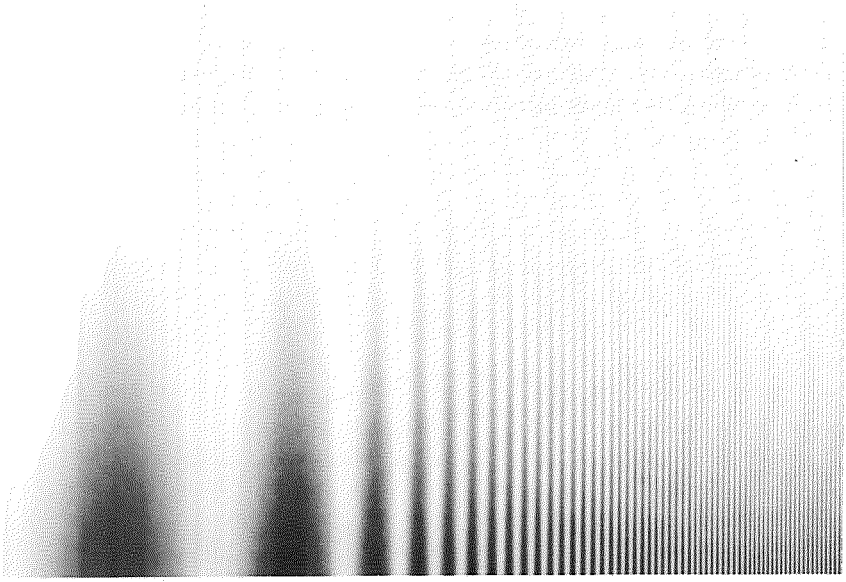


FIGURE 2.26—*Sinusoidally modulated pattern.*

modulation transfer function of the visual system. The visual system enhances edges at abrupt changes in intensity. Each block in the gray scale in figure 2.25 has uniform intensity. However, each block appears to be darker near its lighter neighbor and lighter near its darker neighbor. This apparent overshoot in brightness is a consequence of the spatial frequency characteristic of the eye [43].

2.8.2 Color

Varying the wavelength of light that produces a visual stimulus changes the perceived color from violet (shortest visible wavelength), through blue, green, yellow, and orange, to red. Although the eye can simultaneously discriminate 20 to 30 gray levels, it has the ability to distinguish a larger number of colors [44].

Color can be described by the attributes hue, saturation, and brightness. Any color can be generated by the superposition of three primary colors (usually red, green, and blue) in an additive system or by the subtraction of three primaries (cyan, magenta, and yellow) from white in a subtractive system. Color pictures are produced from digital multiimages by selecting three components for modulation of the primary colors. The brightness values in these three component images are called tristimulus values.

Color order systems based on the principle of equal visual perception of small color differences are of interest for assessing color characteristics for visual perception [45]. In the Munsell system, consisting of a cylindrical space (fig. 2.27), hue is represented by the polar angle; saturation, by the radius; and brightness, by the distance on the cylinder axis (achromatic axis). All gray shades lie on the achromatic axis, because black-and-white images have no saturation or hue. Manipulation of color

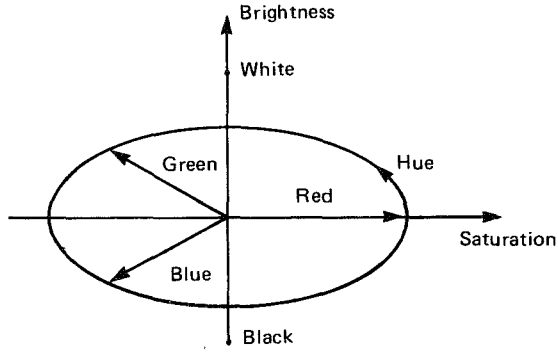


FIGURE 2.27—*Color-perception space.*

images in this space (e.g., filtering) is possible without uncontrolled influence on the relative color balance. (See sec. 4.4.)

2.8.3 Texture

An object or pattern is not perceived by the visual system as an array of independent picture elements. Rather, objects are usually seen as spatially coherent regions on a background. The spatial structure of subpatterns in these regions, characterized by their brightness, color, size, and shape, describes texture. The local subpattern properties give rise to such characteristics of the texture as the perceived lightness, the coarseness, and the directionality. Texture may be defined in terms of regular repetitions of subpatterns [46], or in terms of the frequency of occurrence of averages of local image properties [47, 48].

REFERENCES

[1] Lowe, D. S.: Nonphotographic Optical Sensors, *in* Lintz, J.; and Simonett, D. S., eds.: Remote Sensing of Environment. Addison-Wesley, Reading, Mass., 1976, pp. 155–193.

[2] Stockham, T. S.: Image Processing in the Context of a Visual Model, Proc. IEEE, vol. 60, 1972, pp. 828–842.

[3] Fraser, R. S.; and Curran, R. J.: Effects of the Atmosphere on Remote Sensing, *in* Lintz, J.; and Simonett, D. S., eds.: Remote Sensing of Environment. Addison-Wesley, Reading, Mass., 1976, pp. 34–84.

[4] Papoulis, A.: The Fourier Integral and Its Applications. McGraw-Hill, New York, 1962.

[5] Lighthill, M. J.: Introduction to Fourier Analysis and Generalized Functions. University Press of Cambridge, Cambridge, England, 1960.

[6] Wong, E.: Stochastic Processes in Information and Dynamical Systems. McGraw-Hill, New York, 1971.

[7] Papoulis, A.: Probability, Random Variables, and Stochastic Processes. McGraw-Hill, New York, 1965.

[8] Taylor, A. E.: Introduction to Functional Analysis. John Wiley & Sons, New York, 1958.

[9] Goodman, J. W.: Introduction to Fourier Optics. McGraw-Hill, New York, 1968.

[10] Watanabe, S.: Karhunen-Loéve Expansion and Factor Analysis, Theoretical Remarks and Applications. Transactions of the Fourth Prague Conference on Information Theory, Prague, Czechoslovakia, 1965.

[11] Taylo, J. T.: Digital Filters for Non-Real-Time Data Processing. NASA Report CR-880, 1967.

[12] Selzer, R. H.: Improving Biomedical Image Quality with Computers. NASA/JPL TR 32–1336, Oct. 1968.

[13] Lancos, C.: Discourse on Fourier Series. Hafner Pub. Co., New York, 1966.

[14] Brault, J. W.; and White, O. R.: The Analysis and Restoration of Astronomical Data Via the Fast Fourier Transform, Astron. Astrophys., vol. 13, 1971, pp. 169–189.

[15] Huang, T. S.: Two-Dimensional Windows, IEEE Trans. Audio Electroacoust., vol. AU-20, Mar. 1972, pp. 88–89.

[16] Bracken, P. A.; Dalton, J. T.; Quann, J. J.; and Billingsley, J. B.: AOIPS—An Interactive Image Processing System. National Computer Conference Proceedings, American Federation of Information Societies Press, 1978, pp. 159–171.

[17] Oppenheim, A. V.; Schafer, R. W.; and Stockham, T. G.: Nonlinear Filtering of Multiplied and Convolved Signals, Proc. IEEE, vol. 56, Aug. 1968, pp. 1264–1291.
[18] Andrews, H. C.; and Hunt, R.: Digital Image Restoration. Prentice Hall, Englewood Cliffs, N.J., 1977.
[19] ERTS Data Users Handbook. NASA Doc. 712D4249, Washington, D.C., 1972.
[20] Mikhail, E. M.; and Baker, J. R.: Geometric Aspects in Digital Analysis of Multispectral Scanner Data. American Society of Photogrammetry, Washington, D.C., Mar. 1973.
[21] Kratky, V.: Cartographic Accuracy of ERTS, Photogramm. Eng., vol. 8, 1974, pp. 203–212.
[22] Hufnagel, R. E.; and Stanley, N. R.: Modulation Transfer Function Associated with Image Transmission through Turbulent Media, J. Opt. Soc. Am., vol. 54, 1964, pp. 52–61.
[23] Peterson, D. P.; and Middleton, D.: Sampling and Reconstruction of Wave-Number-Limited Functions in *N*-Dimensional Spaces, Inf. Control, vol. 5, 1962, pp. 279–323.
[24] Bracewell, R. N.: The Fourier Transform and Its Applications. McGraw-Hill, New York, 1965.
[25] Legault, R.: The Aliasing Problem in Two-Dimensional Sampled Imagery, *in* Biberman, L. M., ed.: Perception of Displayed Information. Plenum Press, New York, 1973.
[26] Hunt, B. R.; and Breedlove, J. R.: Scan and Display Considerations in Processing Images by Digital Computer, IEEE Trans. Comput., vol. C-24, 1975, pp. 848–853.
[27] Gaven, J. V.; Taritian, J.; and Harabedian, A.: The Informative Value of Sampled Images as a Function of the Number of Gray Levels Used in Encoding the Images, Photogr. Sci. Eng., vol. 14, 1970, pp. 16–20.
[28] Wood, R. C.: On Optimum Quantization, IEEE Trans. Inf. Theory, vol. IT-15, 1969, pp. 248–252.
[29] Huang, T. S.: PCM Picture Transmission, IEEE Spectrum, vol. 2, no. 12, 1965, pp. 57–60.
[30] Andrews, H. C.: Computer Techniques in Image Processing. Academic Press, New York, 1970.
[31] Brigham, E. O.: The Fast Fourier Transform. Prentice Hall, Englewood Cliffs, N.J., 1974.
[32] Cooley, J. W.; and Tukey, J. W.: An Algorithm for the Machine Calculation of Complex Fourier Series, Math. Comput., vol. 19, 1965, pp. 297–301.
[33] Cooley, J. W.; Lewis, A. W.; and Welch, P. D.: Application of the Fast Fourier Transform to Computation of Fourier Integrals, Fourier Series, and Convolution Integrals, IEEE Trans. Audio Electroacoust., vol. AU-15, 1967, pp. 79–84.
[34] Eklundh, J. O.: A Fast Computer Method for Matrix Transposing, IEEE Trans. Comput., vol. C-21, 1972, pp. 801–803.
[35] Rindfleisch, T.: JPL Communication, 1971.
[36] Ahmed, N.; Natarajan, T.; and Rao, K. R.: Discrete Cosine Transform, IEEE Trans. Comput., vol. C-23, 1974, pp. 90–93.
[37] Hunt, B. R.: A Matrix Theory Proof of the Discrete Convolution Theorem, IEEE Trans. Audio Electroacoust., vol. AU-19, 1971, pp. 285–288.
[38] Oppenheim, A. V.; and Schafer, R. W.: Digital Signal Processing. Prentice Hall, Englewood Cliffs, N.J., 1975.
[39] Hunt, B. R.: Digital Image Processing, Proc. IEEE, vol. 63, 1975, pp. 693–708.
[40] Stockham, T. G.: The Role of Psychophysics in the Mathematics of Image Science. Symposium on Image Science Mathematics, Monterey, Calif., Western Periodicals Comp., Nov. 1976, pp. 57–59.
[41] Jacobson, H.: The Information Capacity of the Human Eye, Science, vol. 113, Mar. 1951, pp. 292–293.
[42] Lipkin, B. S.: Psychopictorics and Pattern Recognition, SPIE J., vol. 8, 1970, pp. 126–138.

[43] Cornsweet, T. N.: Visual Perception. Academic Press, New York and London, 1970.
[44] Sheppard, J. J.; Stratton, R. H.; and Gazley, C. G.: Pseudo-Color as a Means of Image Enhancement, Am. J. Optom., vol. 46, 1969, pp. 735–754.
[45] Billmeyer, F. W.; and Saltzmann, M.: Principles of Color Technology. Interscience, New York, 1966.
[46] Hawkins, J. K.: Textural Properties for Pattern Recognition, *in* Lipkin, B. S.; and Rosenfeld, A.: Picture Processing and Psychopictorics. Academic Press, New York and London, 1970.
[47] Rosenfeld, A.: Visual Texture Analysis: An Overview. TR-406 (University of Maryland, College Park, Md.), Aug. 1975.
[48] Haralick, R. M.; Shanmugam, K.; and Dirnstein, I.: Texture Features for Image Classification, IEEE Trans. Systems, Man Cybernetics, vol. SMC-3, 1973, pp. 610–621.

Page intentionally left blank

3. Image Restoration

3.1 Introduction

Image restoration is concerned with the correction for distortions, degradations, and noise induced in the imaging process. The problem of image restoration is to determine a corrected image $\hat{f}(x', y')$ from the degraded recorded image $g(x, y)$ that is as close as possible, both geometrically and radiometrically, to the original object radiant energy distribution $f(x', y')$.

In section 2.3 a linearized model for the imaging process that separated geometric and radiometric degradations was developed:

$$g = T_G T_R f + n_s \tag{2.81}$$

The geometric distortions T_G are represented by the coordinate transformation, (2.76). The radiometric degradation T_R, equation (2.80), is given by the convolution of the object radiant energy with the system point spread function (PSF) h and by random noise n_r. The additive term n_s includes coherent noise, temperature effects on detector response, and camera-dependent errors.

Formally, an estimate $\hat{f}$ of the original scene f is obtained by

$$\hat{f} = T_R^{-1} T_G^{-1} (g - n_s) \tag{3.1}$$

The effects of the atmosphere and n_s are removed by preprocessing, described in the next section. Geometric corrections T_G^{-1} and radiometric restoration T_R^{-1} are discussed in sections 3.3 and 3.4, respectively.

3.2 Preprocessing

The purpose of preprocessing is to remove the atmospheric effects described by equation (2.77) and the degradations and noise represented by the term n_s in equation (2.81). Image analysis is susceptible to changes in illumination, in atmospheric conditions, in Sun angle, in viewing angle, and in surface reflectance and to systematic instrument errors. Some of these effects are greater in aircraft multispectral data than in satellite images. Preprocessing should remove all systematic variations from the data so that the effects of signal changes are minimized. Preprocessing uses a priori information. For example, Landsat Return Beam Vidicon (RBV) images have to be corrected for nonuniform response of the

vidicon surface (shading), and Landsat Multispectral Scanner (MSS) images require correction for variations of detector gain and offset.

Illumination and atmospheric effects are also removed by preprocessing. After removal of the path radiance (see eq. (2.77)), multiplicative effects that are correlated between channels of multispectral images can be reduced by ratioing pairs of data channels (see sec. 4.5.2). For classification (see ch. 8), atmospheric differences between training areas and areas to be classified can cause changes in both magnitude and spectral distribution of signals and consequently misclassifications. Because preprocessing permits the use of multispectral pixels or signatures from localized areas to be applied to other locations and conditions, preprocessing techniques are frequently called signature extension techniques.

3.2.1 Illumination Correction

The correction for different lighting conditions in remote sensing images due to variable solar elevation cycles is important for comparison of the reflection of materials in different areas and for the generation of mosaics of images taken at different times.

A first-order correction ignores topographical effects and the dependence of the back scattering term in equation (2.77) on the Sun angle. This correction adjusts the average brightness of frames, and consists of a multiplication of each pixel with a constant derived from the Sun elevation angle. For Landsat MSS images the Sun angle effect is also dependent on the latitude [1].

3.2.2 Atmospheric Correction

Atmospheric effects in remotely sensed images are primarily due to atmospheric attenuation of radiation emanating from the surface, to Rayleigh and aerosol scattering of solar radiation between the sensor and the scene, and to sensor scan geometry. (See eq. 2.1.) Scattering is the most serious effect. Aerosol scattering produces a luminosity in the atmosphere often called haze. Scattering is wavelength dependent, where the shorter wavelengths are most affected. In multispectral images the blue and green components often have visibly less scene contrast than the red and infrared.

For Landsat MSS data, path radiance constitutes a major part of the received signal in band 4 and is not negligible in band 7 [1]. There is also a strong dependence of the radiance on haze content and some dependence on the scan angle, even though the MSS scans only about 6° from the nadir. A crude first-order correction for the path radiance is based on the assumption that areas with zero reflectance can be located in a spectral component of a multi-image [2]. The reflectance of water is

essentially zero in the near-infrared region of the spectrum such as in band 7 of the Landsat MSS. Therefore, it can be assumed that the signal of clear open water represents the path radiance. The histograms of the other spectral components for the same area are plotted. The lowest pixel value in each component is used as an estimate for the path radiance and is subtracted from each pixel.

The sensor scan geometry and its relationship to the Sun position are important factors in remote sensing from aircraft and satellites. The longer observation path through the atmosphere for larger scan angles tends to reduce the received signal. An opposite effect is caused by scattering in the atmosphere, and this scattering adds extraneous path radiance to the signal. The relative balance between the two effects depends on the direction of the scan in relation to the Sun position [1]. Techniques for correcting remotely sensed data for Sun angle and atmospheric effects are described in [1–6].

As the state of the science of remote sensing advances, it becomes increasingly necessary to compare data obtained at different times and by different sensors. This comparison requires the determination of absolute target reflectances. The process of finding the correspondence between measurements and a quantity in a system of units is called calibration [7, 8].

3.2.3 Noise Removal

The removal of coherent instrument and transmission noise is important, so that subsequent image enhancement, image registration, and numerical image analysis can be performed on images with a high signal-to-noise ratio. The precise separation of any noise from the data must be based on quantifiable characteristics of the noise signal that distinguish it uniquely from the other image components. The essence of coherent noise removal is to isolate and remove the identifiable and characterizable noise components in a manner that does a minimum of damage to the actual image data. In most cases, the errors caused in the real signal by the removal process, although small, vary from point to point and can only be measured if detailed knowledge about the scene is available. The main types of coherent noise appearing in images are periodic, striping, and spike noises.

Periodic noise may be caused by the coupling of periodic signals related to the raster-scan and data-sampling mechanism into the imaging electronics of electro-optical scanners or by power consumption variations and mechanical oscillations in electromechanical scanners or tape recorders. The recorded images contain periodic interference patterns, with varying amplitude, frequency, and phase, superimposed on the original scene. For typical spacecraft systems, these periodic noises often exhibit phase coherence over times that are long compared to the frame

time of the camera. For this reason, the periodic noise appears as a two-dimensional pattern exhibiting periodicity along the scan lines and perpendicular to them. This periodicity is characterized in the two-dimensional Fourier domain, where the coherent noise structure appears in the two-dimensional amplitude spectrum as a series of spikes, representing energy concentrations at specific spatial frequency locations. The removal of periodic noise components can be achieved by band-pass or notch filtering. (See sec. 2.2.5.)

Figure 3.1*a* shows an image obtained by an electromechanical aircraft scanner. The image is degraded by a strong periodic interference pattern and by data dropouts visible as black spots. The magnitude of the Fourier transform is shown in figure 3.1*b*.

The spikes parallel and under a slight angle to the vertical frequency axis represent periodic noise components. This effect can be demonstrated by observing a single periodic component. Let $n_1(x, y)$ be a two-dimensional sinusoidal pattern with spatial frequencies u_0, v_0 and with amplitude A:

$$n_1(x, y) = A \cos 2\pi(u_0 x + v_0 y) \tag{3.2}$$

The Fourier transform of $n_1(x, y)$ is [9]

$$N_1(u, v) = \frac{A}{2}[\delta(u-u_0, v-v_0) + \delta(u+u_0, v+v_0)] \tag{3.3}$$

and it represents a pair of impulses at (u_0, v_0) and $(-u_0, -v_0)$ in the spatial frequency plane. The line connecting the two impulses is perpendicular to the cosine wave. The Fourier spectrum in figure 3.1*b* indicates that the noise is composed of several periodic components. The noise components along two lines parallel to the vertical frequency axis may be due to scan-line-dependent random phase shifts in a horizontal periodic noise pattern with frequency u_0:

$$n_2(x, y) = B \cos [2\pi u_0 + \phi(y)] \tag{3.4}$$

If the phase $\phi(y)$ is assumed to be linearly dependent on the scan location—$\phi(y) = cy$—the Fourier transform of $n_2(x, y)$ is

$$N_2(u, v) = \frac{B}{2}[\delta(u-u_0, v-c) + \delta(u+u_0, v+c)] \tag{3.5}$$

Thus, with c varying, $N_2(u, v)$ represents impulses located along two lines parallel to the vertical frequency axis. (See fig. 3.2.)

The effect of removing the noise components with a notch filter is shown in figure 3.3, where part *a* shows the Fourier spectrum after removal of the noise frequencies, and part *b* is the reconstructed image. The periodic noise is not completely removed, and too much of the image information may be affected by this crude filtering procedure. A technique

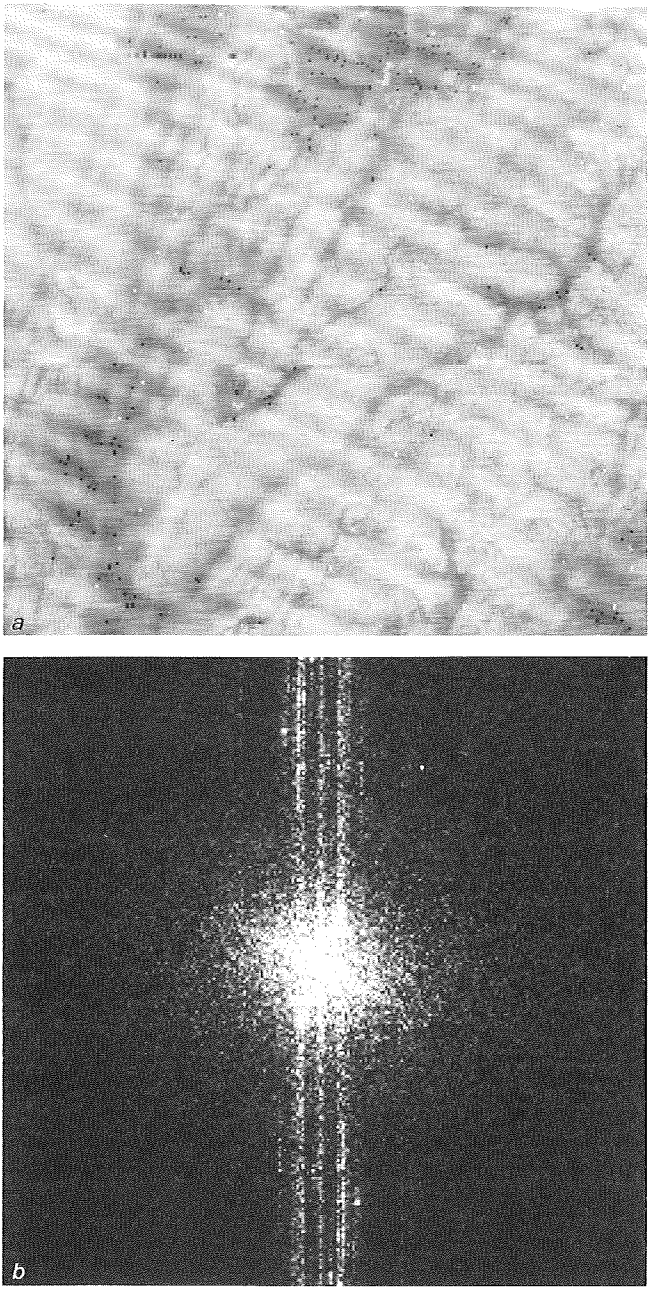


FIGURE 3.1—*Example of periodic noise. (a) Image with periodic interference pattern and spike noise. (b) Magnitude of Fourier transform, showing periodic noise spikes.*

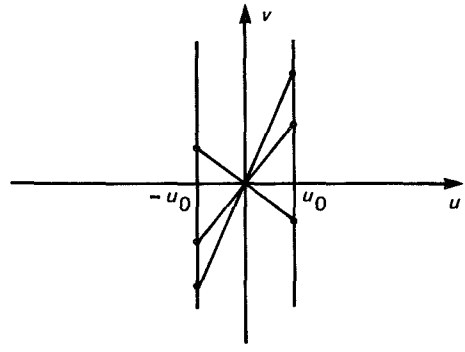


FIGURE 3.2—*Locations of phase-dependent noise components in frequency domain.*

described by Seidman [10] first extracts the principal noise components from the Fourier transforms, creates a noise pattern by the inverse Fourier transform, and subtracts a weighted portion of the noise pattern from the image. Let $G(u, v)$ be the Fourier transform of the noisy image $g(x, y)$. A two-dimensional filter $H(u, v)$ is constructed to pass only noise components such that the noise spectrum is

$$N(u, v) = G(u, v)H(u, v) \tag{3.6}$$

The determination of H requires much judgment and is best performed interactively. Figure 3.4*a* shows the Fourier transform of figure 3.1*a* in the format generated by the discrete Fourier transform (DFT). (See fig. 2.22.) Figure 3.4*b* shows the isolated noise components extracted by the two-dimensional band-pass filter $H(u, v)$. The spatial representation of the noise is obtained by

$$n(x, y) = \mathcal{F}^{-1}\{N(u, v)\} \tag{3.7}$$

To minimize the effect of components not present in the noise estimate n, a weighted portion of n is subtracted from g to obtain an estimate $\hat{f}$ of the corrected image:

$$\hat{f}(x, y) = g(x, y) - w(x, y)n(x, y) \tag{3.8}$$

The weighting function w is determined such that the variance of $\hat{f}$ is minimized over a neighborhood of every point (x, y). Figure 3.5 shows the corrected image $\hat{f}$ obtained by equation (3.8) for a 15- by 15-point spatial neighborhood. Furthermore, the spike noise was removed by the technique described later in this section. The actual noise shown in figure 3.6 is obtained by

$$n_a(x, y) = g(x, y) - \hat{f}(x, y) \tag{3.9}$$

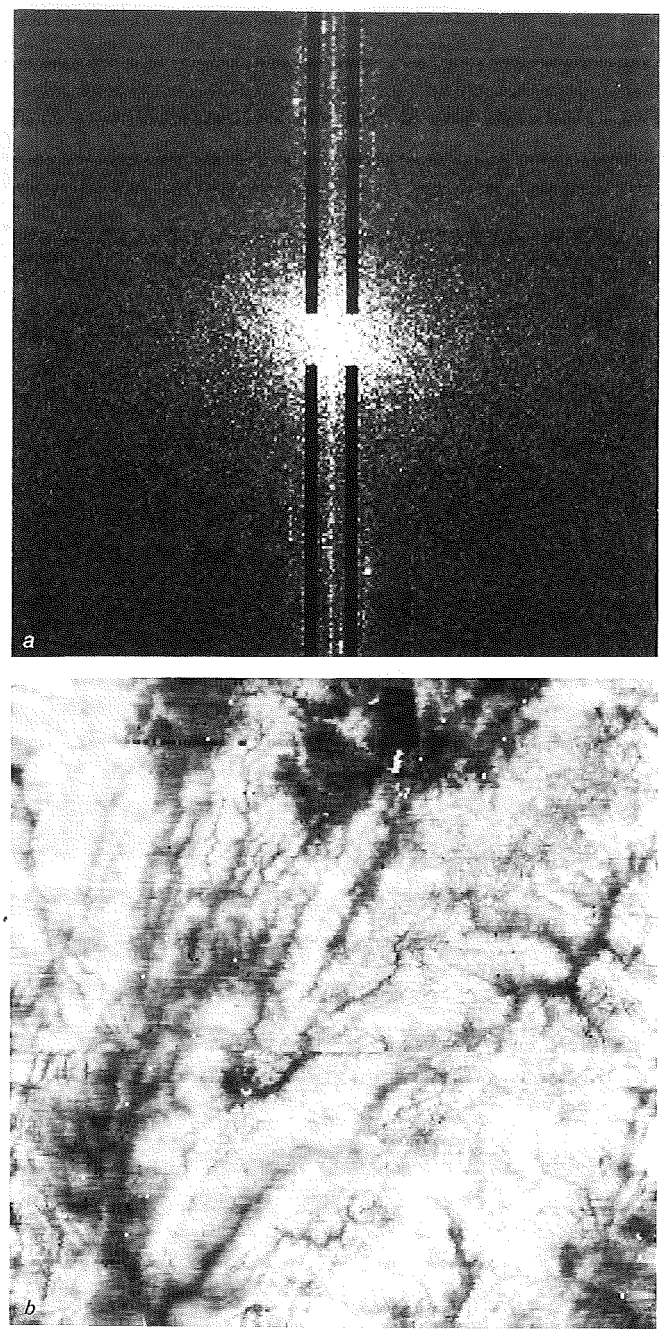


FIGURE 3.3—*Application of frequency-domain filter. (a) Filtered Fourier transform. (b) Corrected image reconstructed from part a.*

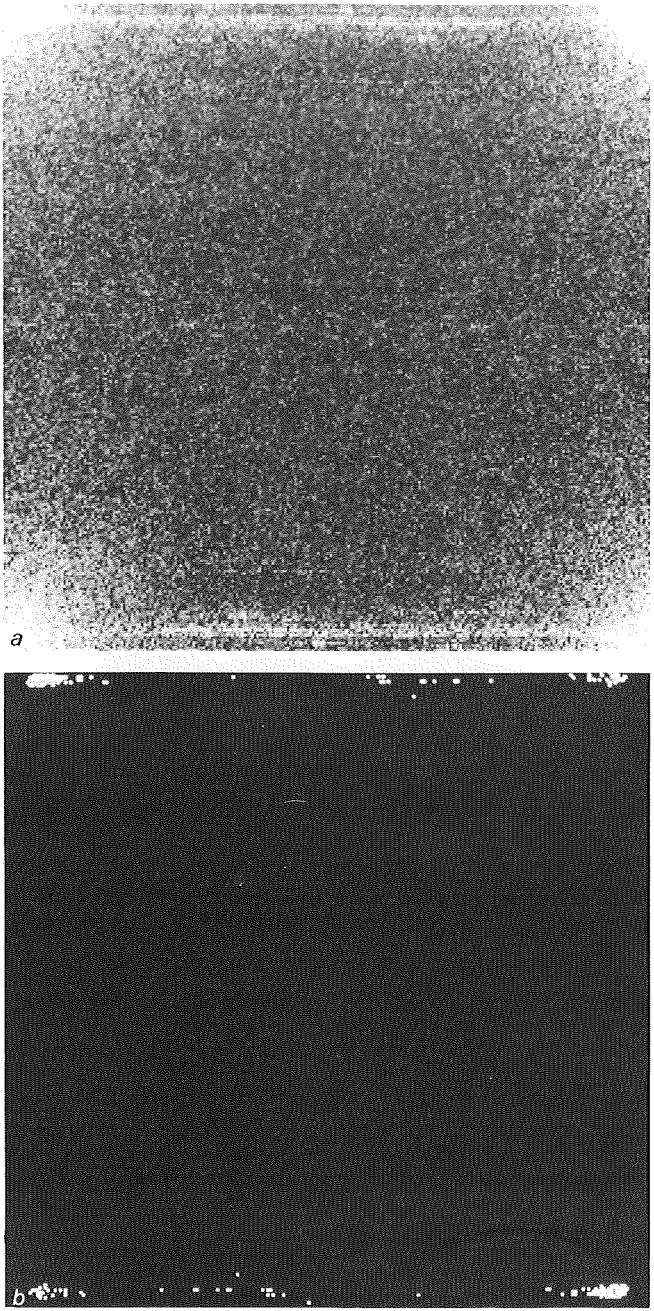


FIGURE 3.4—*Noise filter design. (a) Fourier transform of noisy image. (b) Noise components abstracted by filter.*

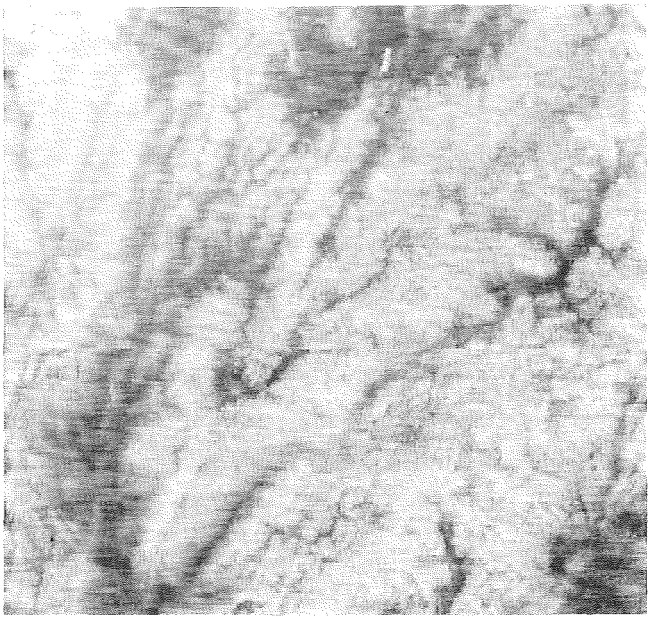


FIGURE 3.5—*Image after removal of periodic and spike noise.*

Striping or streak noise is produced by a variety of mechanisms, such as sensor gain and offset variations, data outages, and tape recorder dropouts. This type of noise becomes apparent as horizontal streaks, especially after removal of periodic noise. The characteristic that distinguishes streak noise from the actual scene is its correlation along the scan line direction and the lack of correlation in the perpendicular direction. This distinction is not complete, because linear features are present in some natural scenes and noise removal based on this characteristic may result in major damage to the true signal in regions that contain scene components resembling the noise. Striping, if not removed before image enhancement, is often also enhanced and may become unacceptable in ratio images. (See sec. 4.5.1.)

A technique to correct for streak noise is to compare the local average intensity values of lines adjacent and parallel to the streaks with the average value of the streak itself and to apply a gain factor to account for any differences. A multiplicative rather than an additive correction is applied because the physical origin of the noise is multiplicative (magnetic tape dropouts). This correction is particularly data dependent in its effect, and although providing a global improvement, it may introduce artifacts in the detail [11].

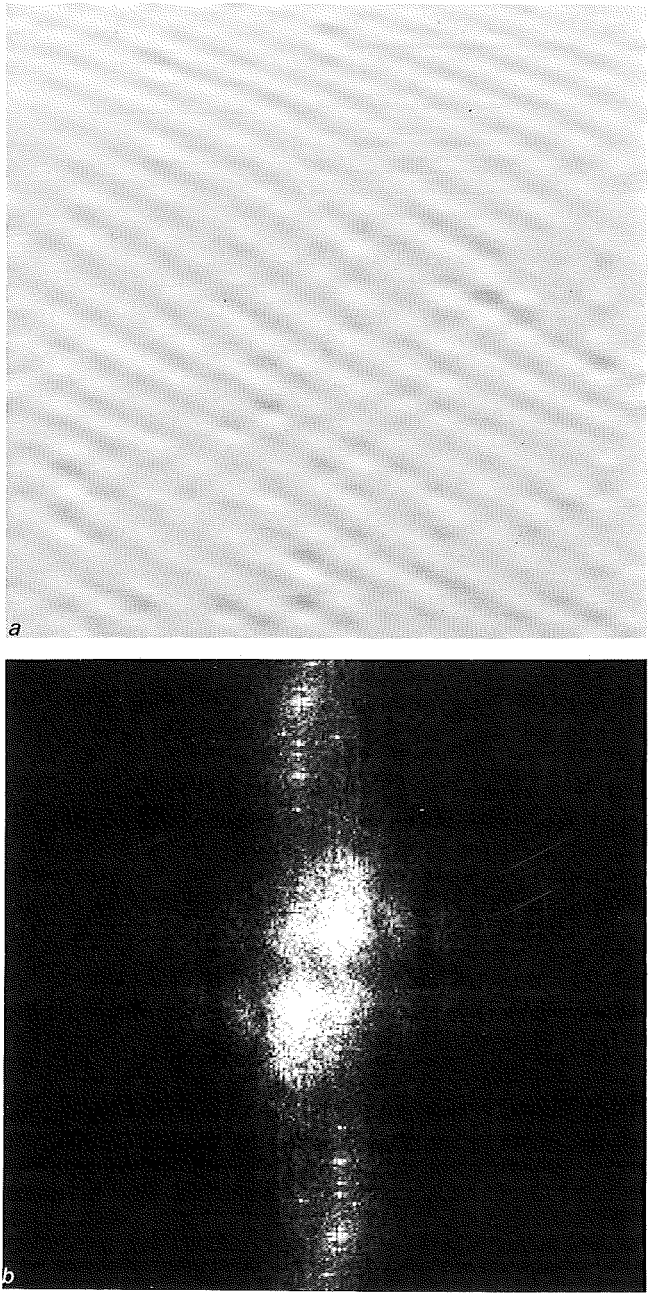


FIGURE 3.6—*Removed noise pattern. (a) Noise subtracted from Figure 3.1a. (b) Magnitude of its Fourier transform.*

Regular striping may occur in images taken by multidetector sensors. The mirror speed of optical–mechanical scanners is limited. For example, the orbital velocity of Landsat is such that the satellite moves forward six fields of view during the time needed to scan one image line. Therefore, six detectors are used in the MSS that image six lines for each spectral band during a single sweep of the mirror. The Visible Infrared Spin Scan Radiometer (VISSR) onboard the Synchronous Meteorological Satellite (SMS) has eight detectors per spectral band.

Let D be the number of detectors in a sensor. Each detector records a subimage consisting of every Dth line. The complete image is formed by interlacing these subimages. The transfer functions of the individual detectors are not identical, because of temperature variations and changes in the detector material. Some detectors have a nonlinear transfer characteristic and a response that depends on their exposure history. Because of these effects images with regular striping are obtained. The corrections derived from scanning a gray wedge at the end of each scan line and applied in ground processing do not remove the striping entirely. Therefore, the image data themselves are used to derive a relative correction of the individual detector subimages such that each one is related in the same way to the actual scene radiance. This correction is based on the assumption that over a sufficiently large region W of size M by N, each sensor is exposed to scene radiances with the same probability distribution.

For a linear and time invariant sensor transfer function, a recorded element of the subimage belonging to detector d is given by

$$g_d(j_d, k) = a_d f_d(j_d, k) + b_d \tag{3.10}$$

$$j_d = d, d+D, d+2D, \ldots$$

$$k = 1, \ldots, N$$

If the gain a_d and offset b_d for each detector are known, a corrected detector output can be calculated by

$$\hat{f}_d(j_d, k) = \frac{g_d(j_d, k) - b_d}{a_d} \tag{3.11}$$

Under the assumption that each subimage has the same mean and variance, the gain and offset are given by

$$a_d = \frac{\sigma}{\sigma_d} \tag{3.12}$$

and

$$b_d = m - a_d m_d \tag{3.13}$$

where m_d and σ_d are the mean gray value and the standard deviation, respectively, of the subimage for detector d; and m and σ are the total

mean gray value and standard deviation, respectively, in the reference region W. For a normal distribution of radiance, transformation (3.11) equalizes the probability distribution of each detector subimage to the probability distribution of the total image. Nonlinear sensor effects distort the distribution, and the linear correction, equation (3.11), does not eliminate striping completely. A nonlinear correction, obtained by matching the cumulative histograms of the individual subimages to the cumulative histogram of the total image, successfully reduces striping [12, 13]. Let H be the cumulative histogram of the entire image; i.e., let $H(f)$ be the number of occurrences of detector output values less than or equal to f. Let H_d be the cumulative histogram for detector d. Then $H_d(g)$ is the number of outputs of detector d less than or equal to g. The transfer function $\hat{f}=f(g)$ is obtained by

$$n_d H(f) \leq n H_d(g) < n_d H(f+1) \tag{3.14}$$

where $n=MN$ and $n_d=MN/D$.

Figure 3.7*a* shows an area of a Landsat-2 MSS image with severe striping. The periodic nature of the striping is also evident as spikes in the vertical power spectrum computed for columns of the image and shown in figure 3.7*b*. The image corrected with the linear transformation (3.11) and its one-dimensional vertical power spectrum are shown in figures 3.8*a* and 3.8*b*, respectively. The result of the nonlinear correction, equation (3.14) by matching the detector histograms is shown in figures 3.8*c* and 3.8*d*. The transfer functions for detectors 1 and 5 obtained by matching detector means and standard deviations and by histogram matching are shown in figure 3.9. Figures 3.10*a* and 3.10*c* show an SMS/VISSR image of the Florida peninsula with striping and after removal of striping by matching the means and standard deviations of the eight detectors to the values of the total image.

Periodic striping noise may also be removed by filtering in the Fourier domain. The contribution of periodic noise to the frequency spectrum is concentrated at points determined by the spatial frequencies of the noise harmonics. For example, the frequency spectrum of a periodic line pattern with discrete frequencies is given by points along a line perpendicular to the direction of the noisy lines in the image. (See sec. 2.2.3.1.) Removing the noise coefficients from the spectrum of the image by interpolation or by notch filtering and taking the inverse transform results in a smoothed picture with deleted noise.

Figure 3.11*a* shows the two-dimensional Fourier spectrum of the image in figure 3.7*a*. The spikes on the vertical frequency axis represent the noise frequencies caused by horizontal striping. Figure 3.11*b* is the corrected image, obtained by one-dimensional frequency-domain filtering with a notch filter that sets the Fourier coefficients at the noise frequencies

in figure 3.7*b* to zero. The ringing near the border, caused by the discrete Fourier transform, is clearly visible. The disadvantage of this technique is that the transform size must be a power of 2, that windowing to reduce ringing is necessary, and that in the case of horizontal striping the image has to be rotated before and after filtering.

Spike noise is caused by bit errors in data transmission or by the occurrence of temporary disturbances in the analog electronics. It produces isolated picture elements that significantly deviate from the surrounding data. Spike noise can be removed by comparing each picture element with its neighbors. If all differences exceed a certain threshold, the pixel is considered a noise point and is replaced by the average of its neighbors. The spike noise in figure 3.1*a* was removed with this technique.

Additive random noise in a recorded image g,

$$g=f+n \tag{3.15}$$

may be suppressed by averaging if multiple frames of an invariant scene are available. The average of a given set of L images g_i is obtained by

$$\bar{g}=\frac{1}{L}\sum_{i=1}^{L} g_i \tag{3.16}$$

If the noise n is uncorrelated and has zero mean, then

$$E\{\bar{g}\}=f \tag{3.17}$$

and

$$\sigma_{\bar{g}}^2=\frac{1}{M}\sigma_f^2 \tag{3.18}$$

where $\sigma_{\bar{g}}^2$ is the variance of the average. Thus, $\bar{g}$ will approach the original image f if L is sufficiently large. This technique requires a very accurate registration of the images. Multiple frames taken at the same time are generally not available for remotely sensed images. Therefore, the primary application of this method is in reducing the noise in digitized maps and photographs used as ancillary data for the analysis of remotely sensed data.

3.3 Geometric Transformations

Geometric transformations are used to correct for the geometric distortions T_G (see sec. 2.4); i.e., they are used to perform the inverse transformation T_G^{-1}. They are also required to overlay images on other images or maps; to produce cartographic projections; to reduce or expand the size of pictures; to correct the aspect ratio (ratio between scales in the horizontal and vertical direction); and to rotate, skew, and flip images.

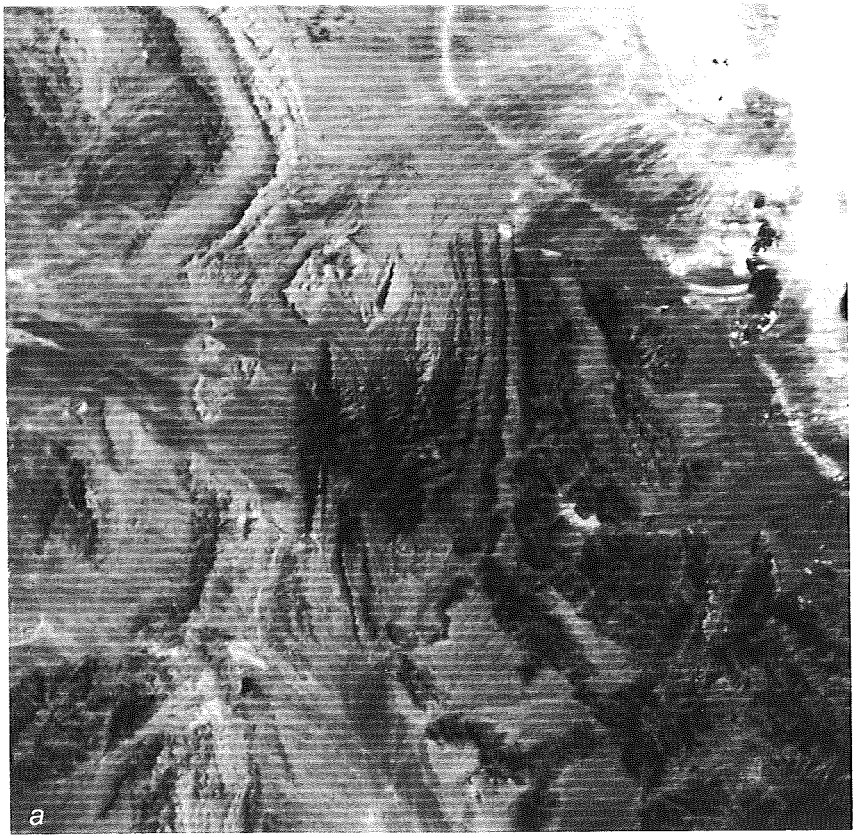


FIGURE 3.7a—*Striped Landsat MSS image with residual striping after radiometric correction at GSFC.*

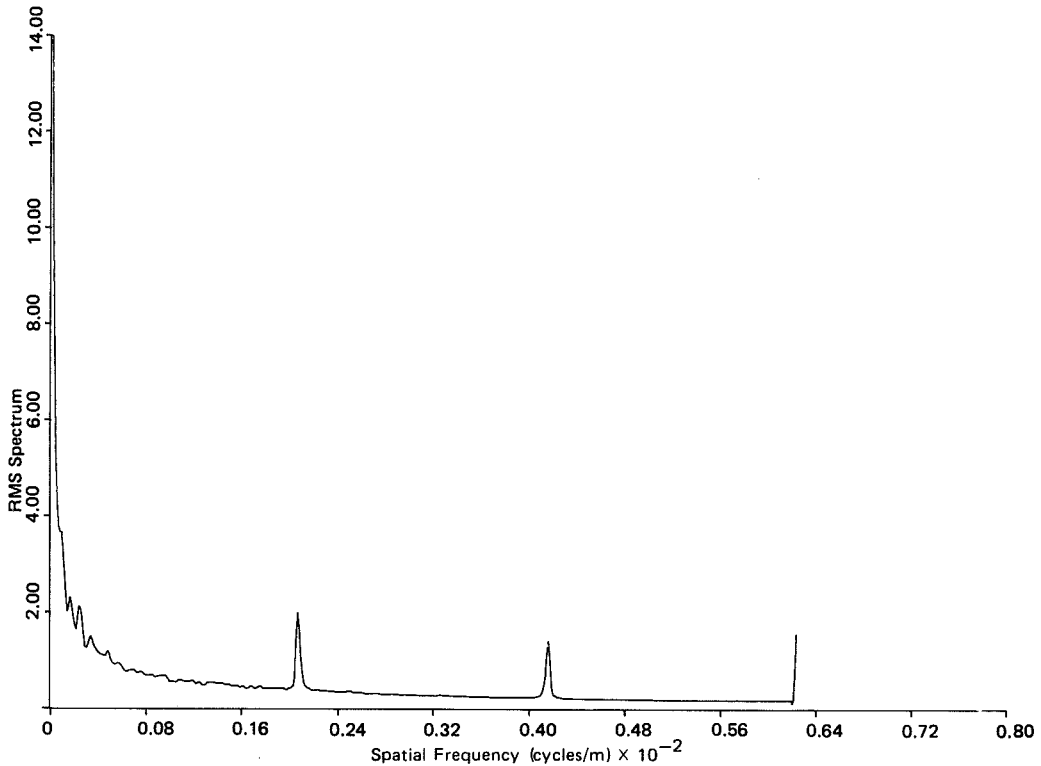


FIGURE 3.7*b*—*One-dimensional vertical power spectrum of image in part a. Striping at frequencies of 0.00211, 0.00422, and 0.00633 cycles per meter (1/6, 1/3, and 1/2 cycles per line).*

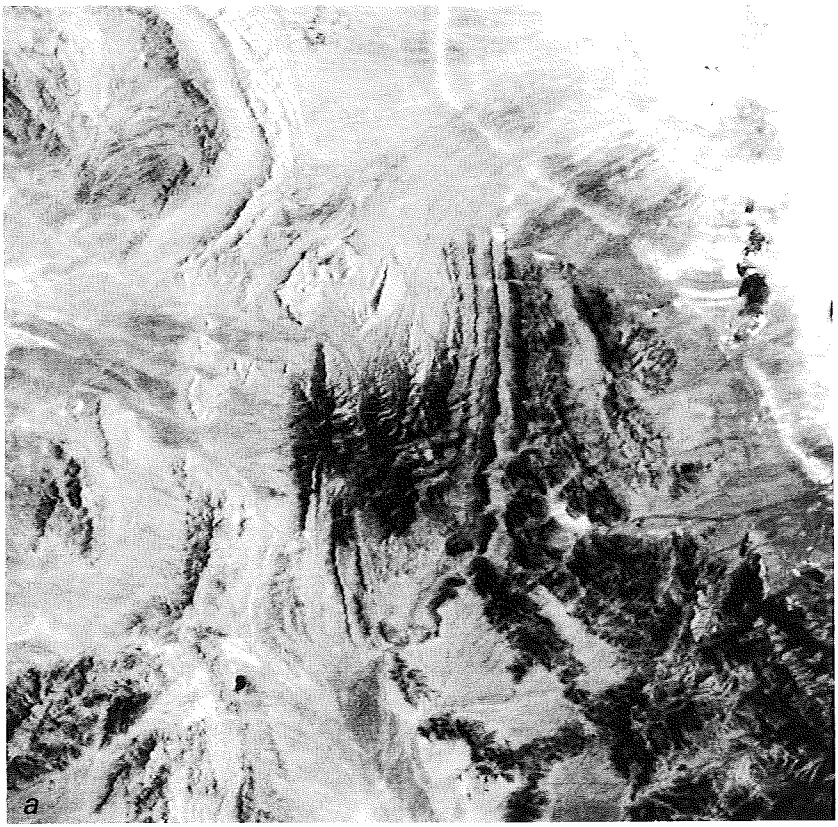


FIGURE 3.8a—*Destriped Landsat MSS image. Striping removed by matching detector means and standard deviations.*

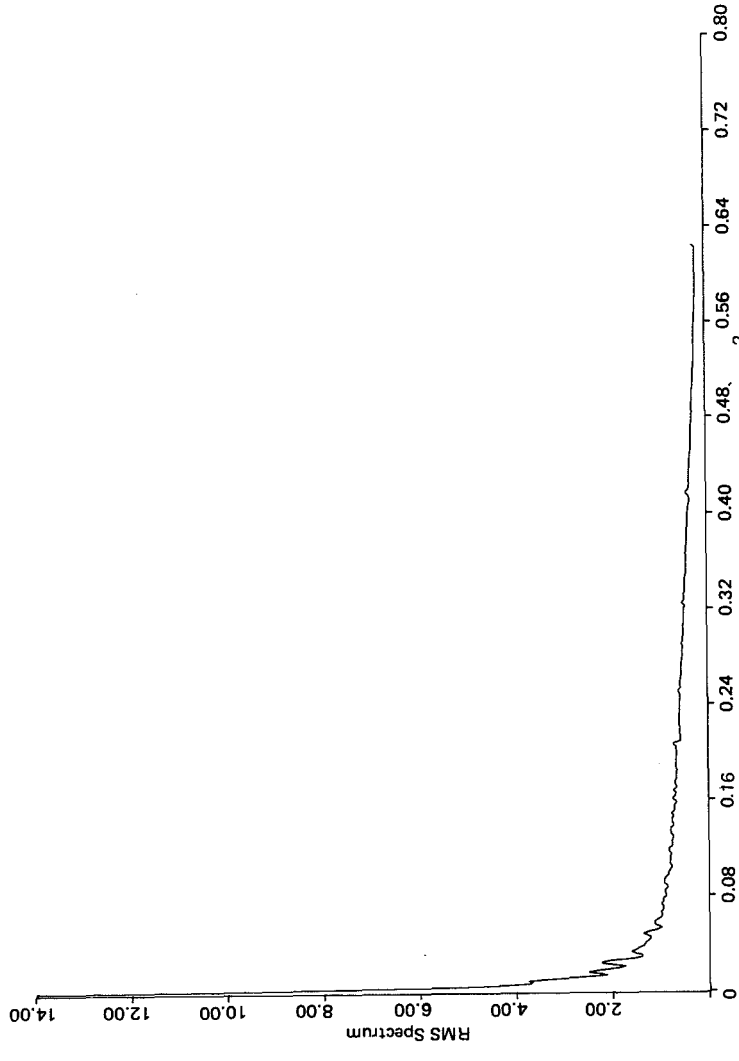


FIGURE 3.8*b*—*One-dimensional vertical power spectrum of image in part a.*

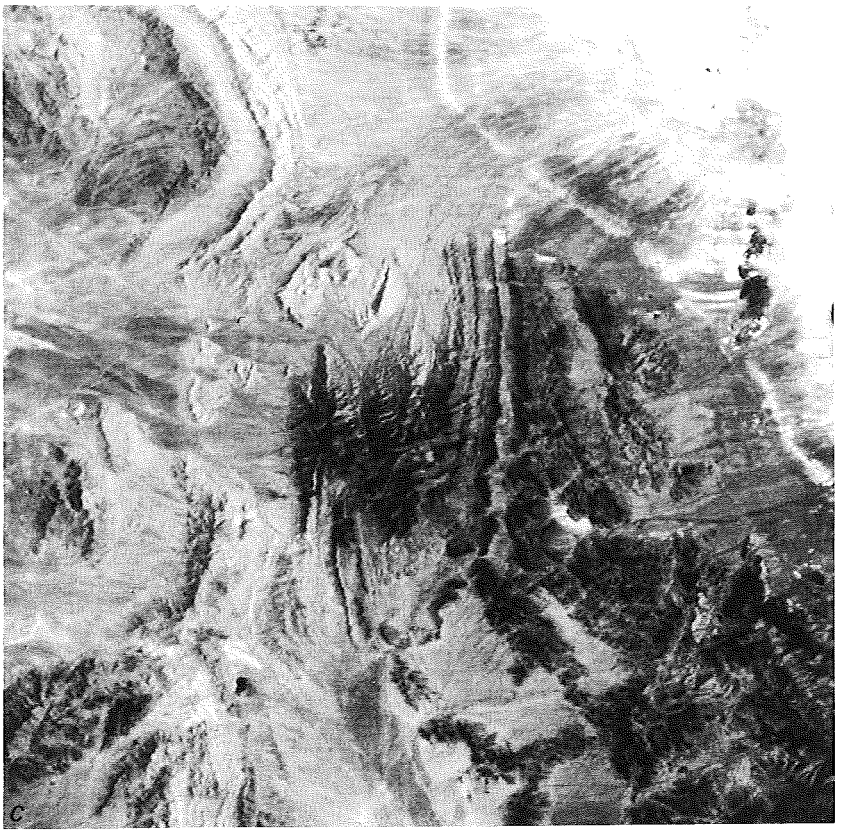


FIGURE 3.8c—*Destriped Landsat MSS image. Striping removed by matching detector histograms.*

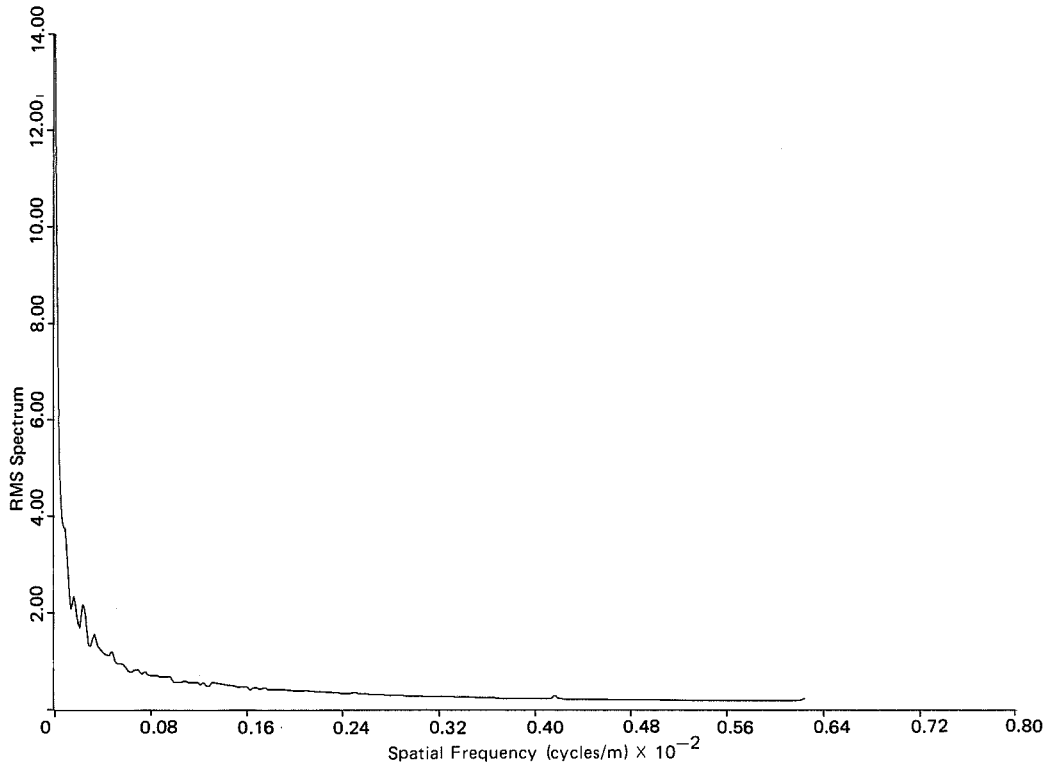


FIGURE 3.8*d*—*One-dimensional vertical power spectrum of image in part c.*

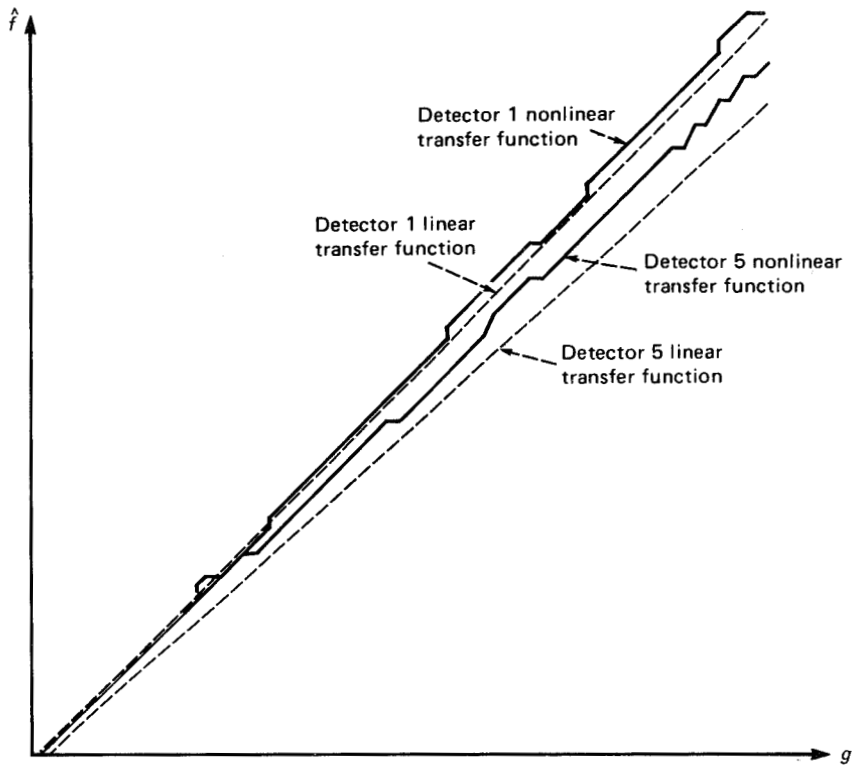


FIGURE 3.9—*Transfer functions for detectors 1 and 5 for correction in figure 3.8c.*

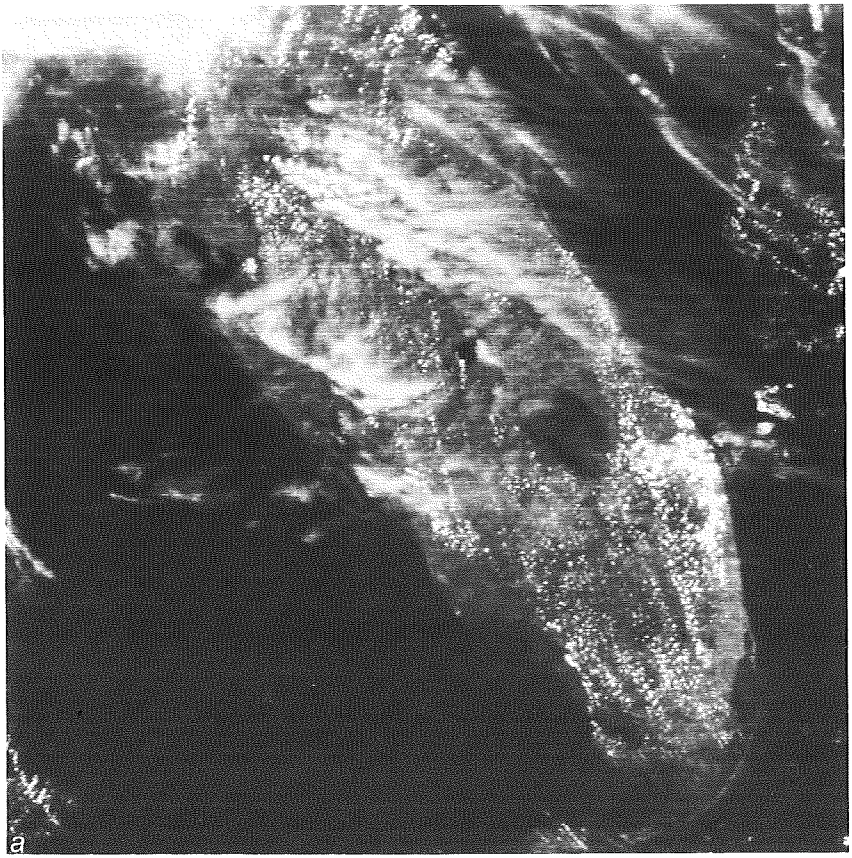


FIGURE 3.10a—*Striped recorded SMS/VISSR image.*

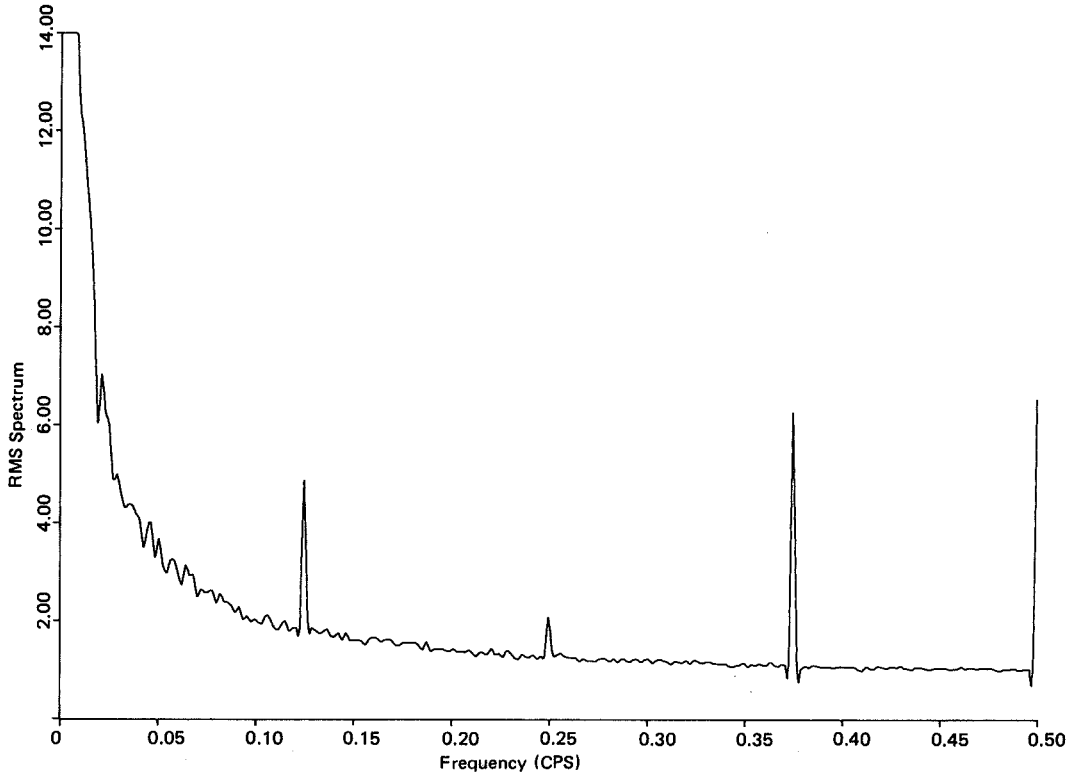


FIGURE 3.10*b*—*Averaged one-dimensional vertical power spectrum of image in part a. Striping occurs at frequencies of 0.125, 0.25, 0.375, and 0.5 cycles per line.*

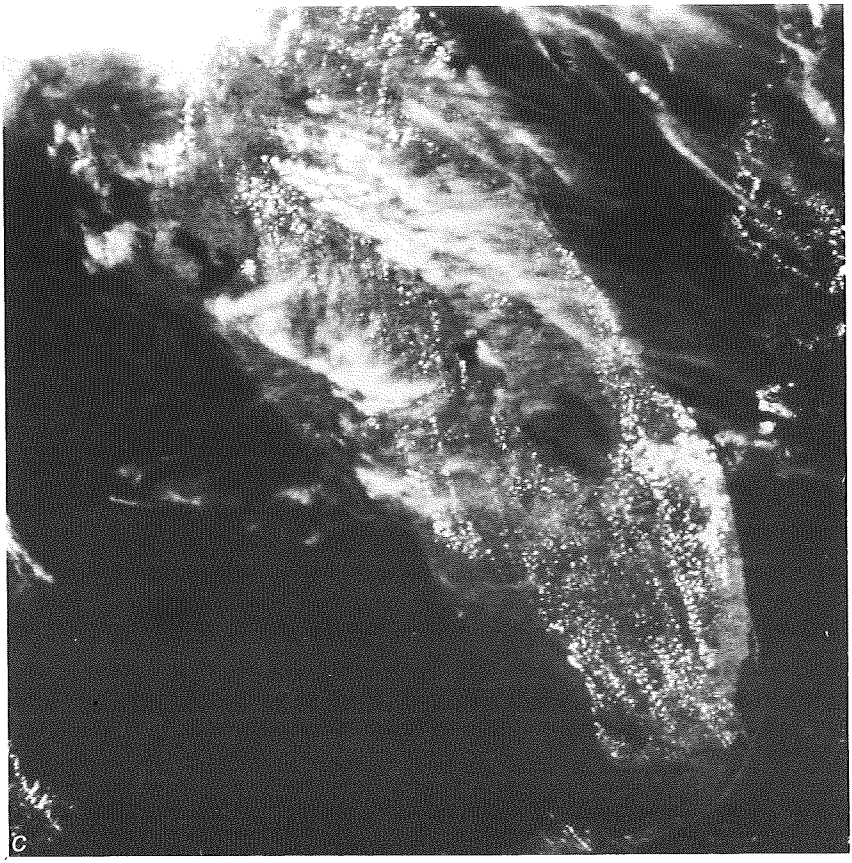


FIGURE 3.10c—*Image in part a after correction to remove striping.*

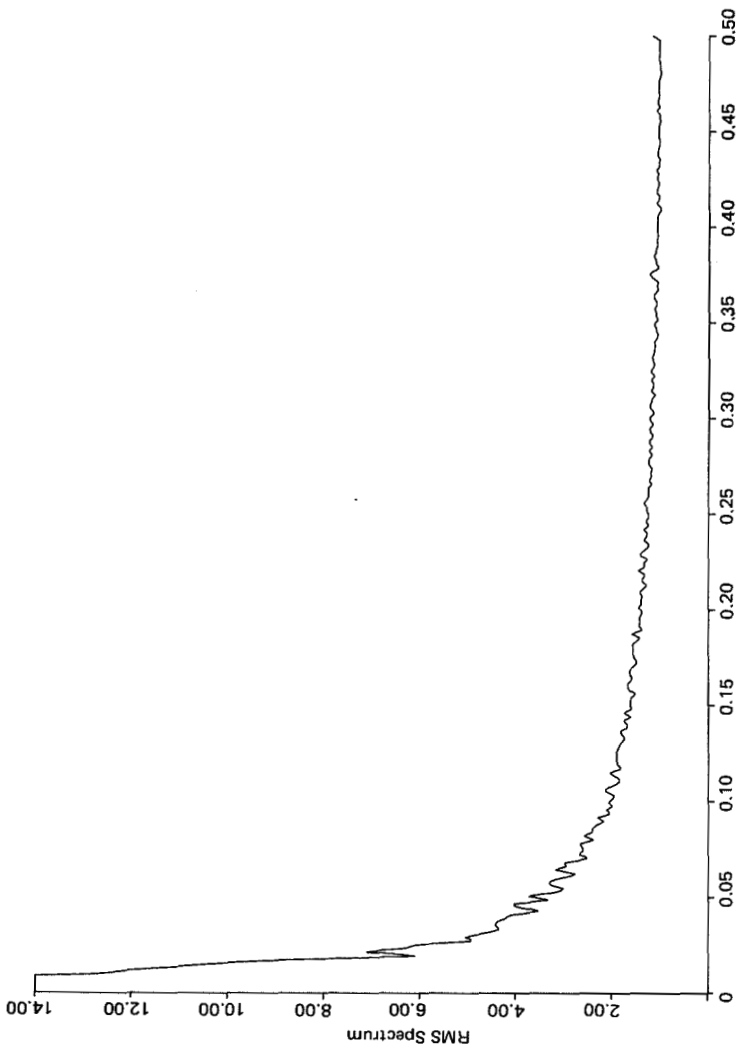


FIGURE 3.10*d*—*One-dimensional vertical power spectrum of image in part c.*

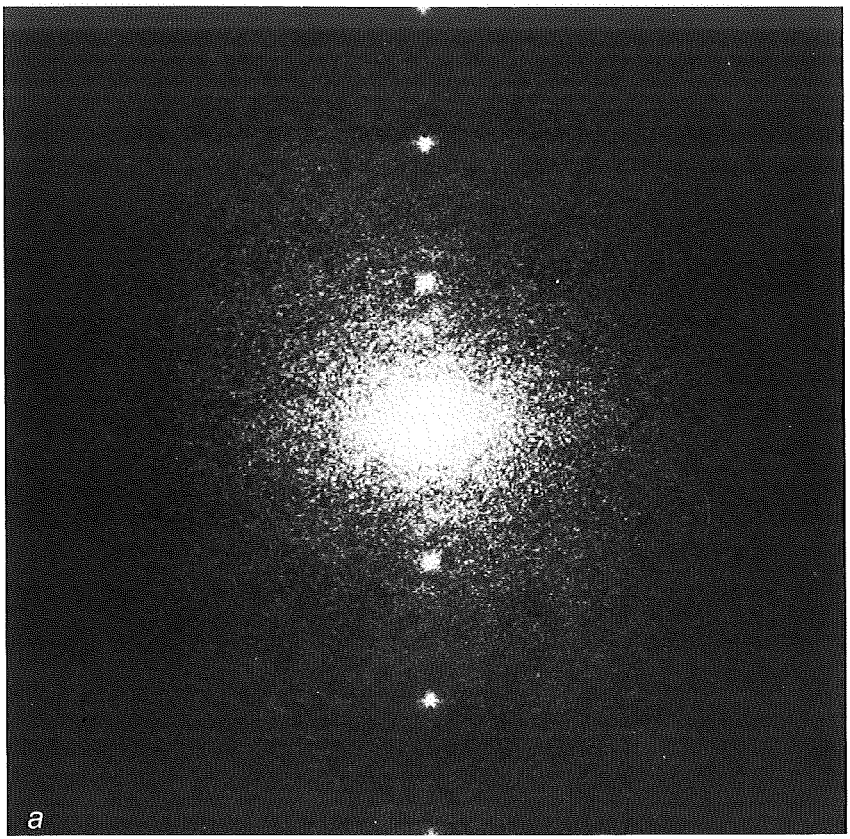


FIGURE 3.11a—*Removal of striping by frequency domain filtering. Two-dimensional Fourier spectrum of image in figure 3.7a.*

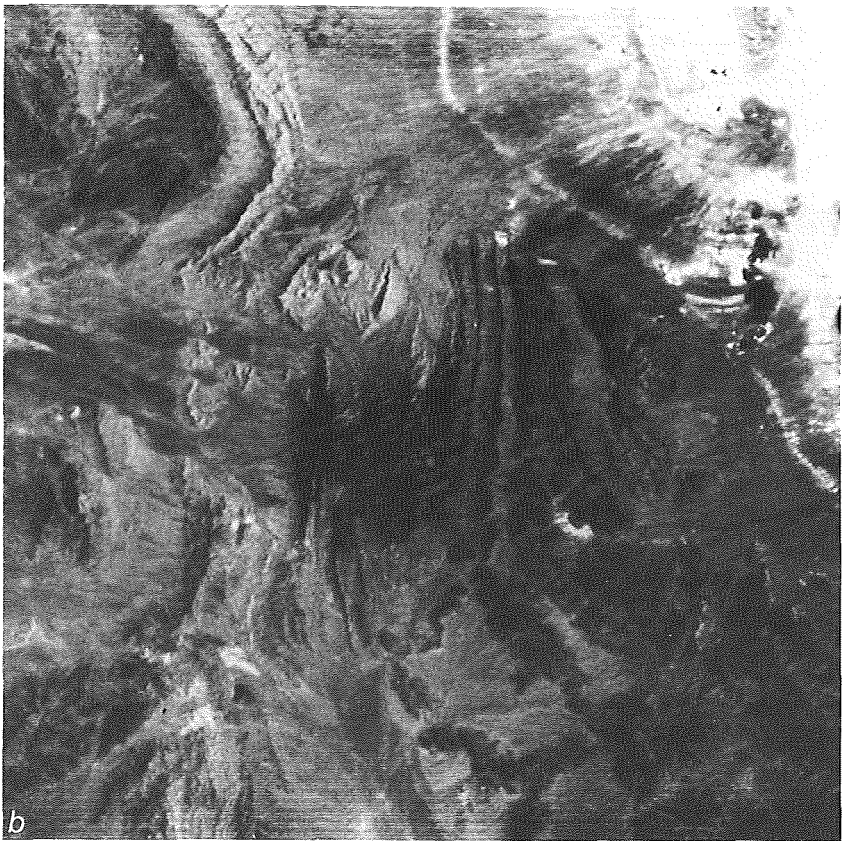


FIGURE 3.11*b*—*Image in figure 3.7a after filtering to suppress striping. Ringing effects near the top and bottom borders are visible.*

3.3.1 Coordinate Transformations

A geometric transformation $\mathbf{T}_g$ is a vector function defined by the equations

$$\left.\begin{aligned} x' &= p(x, y) \\ y' &= q(x, y) \end{aligned}\right\} \tag{2.76}$$

which map a region R in the (x, y) coordinate system into a region R^* in the (x', y') system—$\mathbf{T}_g$: $R \rightarrow R^*$. (See fig. 3.12.) To determine $\mathbf{T}_G{}^{-1}$ for remotely sensed images, the errors contributing to the geometric distortions discussed in section 2.4.1 must be known. The main error sources are variations in aircraft or spacecraft attitude and velocity, Earth rotation during imaging, instrument nonlinearities, Earth curvature, and panoramic distortion.

Two important classes of mappings $\mathbf{T}_g$ for remotely sensed scanner images are (1) the transformations that relate the image coordinates (x, y) to a geodetic coordinate system, and (2) the transformations that relate the coordinate systems of two images. The precise Earth location of scanner measurements must be determined to produce map overlays and cartographic projections of images (see ch. 6), and for the objective analysis of time sequences of images, such as wind speed calculations from cloud displacements (see sec. 7.4). The coordinate transformations between images are employed in the relative registration of several images

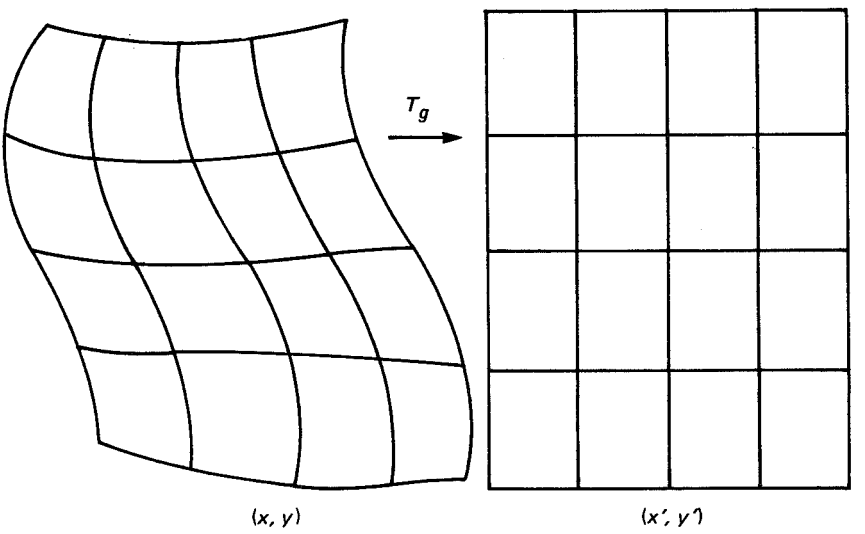


FIGURE 3.12—*Geometric transformation.*

of the same scene (e.g., multispectral, multitemporal, or multisource images). The determination of $\mathbf{T}_g$ for this case is treated in chapter 5.

The derivation of $\mathbf{T}_g$ for the first case is equivalent to the determination of the intersection of the scanner line of sight with the surface of the Earth. Given the aircraft or spacecraft position vector $\mathbf{p}$; the vehicle velocity $\mathbf{v}$; the vehicle attitude in yaw, pitch, and roll; the pointing direction $\mathbf{s}$ of the scanner measured relative to the aircraft or spacecraft axis; and the description of the Earth surface, the location of the intersection $\mathbf{e}$ of the scanner line of sight with the Earth surface can be computed. In the approach described by Puccinelli [14] the Earth coordinate system is chosen as a Cartesian coordinate system (x', y', z') with the origin at the center of the Earth. The spacecraft position vector $\mathbf{p}$ defines the origin of the spacecraft coordinate system about which the attitude changes are measured. The definition of the orientation of the aircraft or spacecraft axes is arbitrary, and different coordinate systems are established for different satellites. For example, the Landsat and Nimbus attitude control systems aline the spacecraft yaw axis with the normal $\mathbf{n}$ from the satellite to the Earth surface. Therefore, the yaw axis is taken to be coincident with the normal vector $\mathbf{n}$; the pitch axis is taken to be $\mathbf{n}\times\mathbf{v}$, which is normal to the orbital plane and to the yaw axis; and the roll axis is $(\mathbf{n}\times\mathbf{v})\times\mathbf{n}$. (See fig. 3.13.) For the spin-stabilized SMS/GOES (Geostationary Operational Environmental Satellite) spacecraft with spin axis $\mathbf{S}$, the yaw axis is taken to be $\mathbf{n}=-\mathbf{p}+(\mathbf{p}\cdot\mathbf{S})\mathbf{S}$, which is perpendicular to $\mathbf{S}$ in the plane defined by $\mathbf{p}$ and $\mathbf{S}$ and generally points to the center of the Earth. The pitch axis is taken to be $\mathbf{S}\times\mathbf{n}$, and the roll axis is the spin axis $\mathbf{S}$ itself [15]. The orientation of the spacecraft axes relative to the Earth coordinate system is given by the orthogonal matrix $\mathbf{D}$, where

$$\mathbf{D}=(\mathbf{c}_1, \mathbf{c}_2, \mathbf{c}_3) \tag{3.19}$$

The column vectors may be chosen as

$$\mathbf{c}_1=\mathbf{n} \qquad \mathbf{c}_2=\frac{\mathbf{n}\times\mathbf{v}}{||\mathbf{v}||^2} \qquad \mathbf{c}_3=\mathbf{c}_2\times\mathbf{c}_1 \tag{3.20}$$

for the first of the previously defined spacecraft coordinate systems. The images coordinates (x, y) of a picture element are determined by the viewing direction $\mathbf{s}$ of the scanner. The direction of the scanner can be described by rotations about the spacecraft yaw, pitch, and roll axes, measured relative to the yaw axis. The three generally time-varying rotation angles define the vector $\mathbf{s}=(\Theta, \Phi, \theta)$, where Θ is the rotation about the yaw axis; Φ is the rotation about the pitch axis; and θ the rotation about the roll axis. Depending on the scanner type, one or two of the rotation angles are zero. For the Landsat MSS, $\Theta=0$, $\Phi=0$, and $\theta=\Delta y(y-y_0)$, where Δy is the angular width of a pixel in the scan

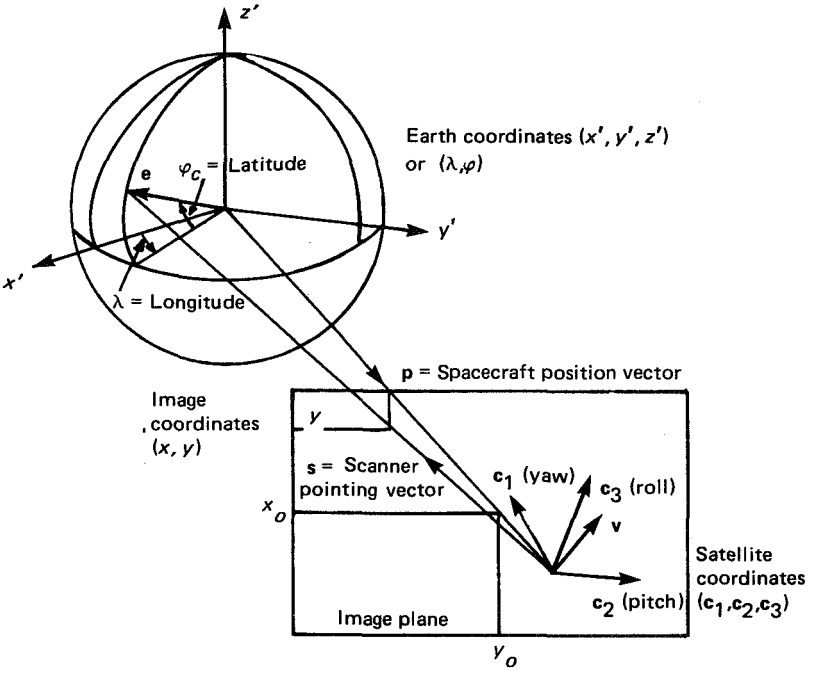


FIGURE 3.13—*Relation between Earth, satellite, and image coordinates.*

direction, and y_0 is the coordinate of the center of the image frame. For the SMS/VISSR, $\Theta=0$, $\Phi=\Delta x(x-x_0)$, and $\theta=\Delta y(y-y_0)$.

The rotations of the spacecraft axis due to yaw τ, pitch ξ, and roll ρ are described by the product of three rotation matrices

$$\mathbf{M}=\begin{pmatrix}\cos\tau & -\sin\tau & 0\\ \sin\tau & \cos\tau & 0\\ 0 & 0 & 1\end{pmatrix}\begin{pmatrix}\cos\xi & 0 & \sin\xi\\ 0 & 1 & 0\\ -\sin\xi & 0 & \cos\xi\end{pmatrix}\begin{pmatrix}1 & 0 & 0\\ 0 & \cos\rho & -\sin\rho\\ 0 & \sin\rho & \cos\rho\end{pmatrix} \tag{3.21}$$

The column vectors of the orthogonal matrix

$$\mathbf{F}=\mathbf{DM} \tag{3.22}$$

are the directions of the spacecraft axis in the Earth coordinate system after a change in the spacecraft attitude. The rotations describing the scanner viewing direction in the spacecraft coordinate system are represented by a matrix $\mathbf{M}'$, identical to $\mathbf{M}$ except that the angles Θ, Φ, and θ replace τ, ξ, and ρ, respectively. To determine the pointing direction of the scanner in the Earth coordinate system, recall that the third column of $\mathbf{F}$ represents the yaw axis in Earth coordinates. Thus, $\mathbf{G}$, where

$$\mathbf{G}=\mathbf{FM}' \tag{3.23}$$

is the orthogonal matrix whose third column is the unit vector representing the scanner line of sight in the Earth coordinate system. Let $\mathbf{m}'$ be the third column of $\mathbf{M}'$; i.e.,

$$\mathbf{m}'=\begin{pmatrix}\cos\Theta\sin\Phi\cos\theta+\sin\Theta\sin\theta\\ \sin\Theta\sin\Phi\cos\theta-\cos\Theta\sin\theta\\ \cos\Phi\cos\theta\end{pmatrix} \tag{3.24}$$

then $\mathbf{g}$, where

$$\mathbf{g}=\mathbf{F}\,\mathbf{m}' \tag{3.25}$$

is the unit vector representing the scanner line of sight in the Earth coordinate system. The Earth surface is defined by the following ellipsoid:

$$\frac{x'^2+y'^2}{a^2}+\frac{z'^2}{c^2}=1 \tag{3.26}$$

The intersection of $\mathbf{g}$ with that surface is given by the vector $\mathbf{e}$, where

$$\mathbf{e}=\mathbf{p}+u\,\mathbf{g} \tag{3.27}$$

The parameter u represents the distance from the scanner to the intersect point and is given by

$$u=\frac{-B-\sqrt{B^2-AC}}{A} \tag{3.28}$$

where

$$\begin{aligned}A&=c^2(g_x^2+g_y^2)+a^2 g_z^2\\ B&=c^2(p_x g_x+p_y g_y)+a^2 p_z g_z\\ C&=c^2(p_x^2+p_y^2)+a^2(p_z^2-c^2)\end{aligned} \tag{3.29}$$

The resulting location vector $\mathbf{e}=(e_{x'}, e_{y'}, e_{z'})$ can be converted to geocentric latitude φ_c and longitude λ:

$$\varphi_c=\tan^{-1}\left(\frac{e_{z'}}{\sqrt{e_{x'}^2+e_{y'}^2}}\right) \tag{3.30}$$

$$\lambda=\tan^{-1}\frac{e_{y'}}{e_{x'}} \tag{3.31}$$

The geodetic map latitude φ is given by

$$\varphi=\tan^{-1}\left(\frac{a^2}{c^2}\right)\tan\varphi_c \tag{3.32}$$

Thus, the geometric transformation relating image coordinates (x, y) to geodetic coordinates (ϕ, λ) is a function of spacecraft position, velocity, attitude, scanner orientation, and surface ellipticity:

$$\mathbf{T}_g=\mathbf{T}_g(\mathbf{p}, \mathbf{v}, \Theta, \phi, \theta, \mathbf{s}, a, c) \tag{3.33}$$

where each positional value in this transformation is in reality a function of time.

The attitude information provided by the satellite instrumentation is generally not accurate enough to meet the precision requirements for geometric correction. To obtain a precision of one picture element for Landsat MSS, each attitude component should be known to 0.1 mrad. However, the Landsat 1 and 2 attitude measurement system is only accurate to 1 mrad [16]. The SMS/GOES VISSR requirements for a precision of one visible picture element demand the determination of the spacecraft attitude to within 5″, or 24 μrad [15].

A precise estimate of the attitude time series over the time interval in which the image was scanned can be obtained by using ground control points. Ground control points are recognizable geographic features or landmarks in the image whose actual locations can be measured in maps. Thus, control points relate an image with the object. The approach is to estimate the attitude time series of the spacecraft with known ground control point locations by least squares [15] or by digital filtering [17].

Pass points are used to determine $\mathbf{T}_g$ for the relative registration of images of the same scene. Pass points are recognizable features that are invariant in a series of images of the same scene but whose absolute locations are unknown. Pass points relate an image with other images while the object coordinates are not known.

In practice the calculation of the exact location of each image point would require a prohibitive amount of computer time. Depending on the nature of the geometric distortions, points in the transformed image may be sparsely and not equally spaced. For reasons of compact data storage and limitations of film recorders, the output picture elements must be regularly spaced. To obtain a continuous picture with equally spaced elements in a reasonable time, the inverse approach is taken. A set of tie points defining a rectangular or quadrilateral interpolation grid in the output image is selected. The exact transformation is only computed for the grid points. The locations of points within each quadrilateral are determined by bilinear interpolation between the vertex coordinates. Values for fractional pixel locations in the input picture are determined by interpolation. The calculation of the location of an output picture element in the original image and interpolation over surrounding pixels is called resampling. (See sec. 3.3.2.)

The actual coordinate transformation can be represented by the two-dimensional polynomials of order m

$$x' = \sum_{j=0}^{m} \sum_{k=0}^{m-j} a_{jk}\, x^j y^k$$
$$y' = \sum_{j=0}^{m} \sum_{k=0}^{m-j} b_{jk}\, x^j y^k \tag{3.34}$$

This transformation is linear in the coefficients a_{jk}, b_{jk}, a property that permits a least-square procedure to be used for determining a_{jk} and b_{jk} from a set of known corresponding tie points (x_i, y_i) and (x_i', y_i').

When the required geometric transformation is of a spatially slowly varying nature, a low-order bivariate polynomial ($m=2$ or 3) usually yields a good approximation of the actual transformation. Second- and third-order polynomials require at least 6 and 10 pairs of corresponding tie points, respectively. It is assumed that the tie points are uniformly distributed throughout the image. Larger differential geometric variations require higher order polynomials to achieve given error bounds. In general, the coefficients associated with the high-order terms in the polynomial are sensitive to the location of the control points. This problem may be avoided by using orthogonal polynomials (e.g., Legendre or Hermite polynomials). Another approach relies on defining low-order polynomials for subareas of the image and assuring continuity at the boundaries between the areas [18]. If the size of the subareas is adapted to the magnitude of the geometric variations, the approximation can be made with arbitrary accuracy, provided that a sufficient number of tie points is available.

A commonly used low-order mapping for a subarea is given by the bilinear transformation

$$\begin{aligned} x' &= a_0 + a_1 x + a_2 y + a_3 xy \\ y' &= b_0 + b_1 x + b_2 y + b_3 xy \end{aligned} \tag{3.35}$$

Four pairs of corresponding tie points are required to determine the coefficients of this transformation. The tie points for the subareas define a net of quadrilaterals. In many cases an affine transformation given by

$$\begin{aligned} x' &= a_0 + a_1 x + a_2 y \\ y' &= b_0 + b_1 x + b_2 y \end{aligned} \tag{3.36}$$

is sufficient to represent the required geometric transformation of a subarea. Affine transformations include rotation, displacement, scaling, and skewing. Three pairs of corresponding tie points are required per subarea defining a net of triangles over the image.

A measure of how the transformation $\mathbf{T}_g$ distorts a coordinate system is given by the Jacobian determinant

$$J = \left| \frac{\partial(x', y')}{\partial(x, y)} \right| = \begin{vmatrix} \dfrac{\partial x'}{\partial x} & \dfrac{\partial x'}{\partial y} \\ \dfrac{\partial y'}{\partial x} & \dfrac{\partial y'}{\partial y} \end{vmatrix} \tag{3.37}$$

For the bilinear transformation (3.35) the Jacobian is given by

$$J = a_1 b_2 - a_2 b_1 + (a_1 b_3 - a_3 b_1) x + (a_3 b_2 - a_2 b_3) y \tag{3.38}$$

For the identity transformation

$$\begin{aligned} x' &= x \\ y' &= y \end{aligned} \tag{3.39}$$

and the coefficients have the following values:

$$a_1 = b_2 = 1,\ a_0 = a_2 = a_3 = b_0 = b_1 = b_3 = 0, \text{ and } J = 1.$$

A singular transformation is characterized by $J = 0$. For the affine transformation (3.36) the Jacobian is

$$J = a_1 b_2 - a_2 b_1 \tag{3.40}$$

Special cases of affine transformations have the following Jacobian determinants:

1. Rotation by angle ϕ:

$$\left.\begin{aligned} x' &= x \cos\phi + y \sin\phi \\ y' &= -x \sin\phi + y \cos\phi \\ J &= 1 \end{aligned}\right\} \tag{3.41}$$

2. Scaling by factors a and b

$$\begin{aligned} x' &= ax \\ y' &= by \\ J &= ab \end{aligned} \tag{3.42}$$

3. Skewing by angle θ

$$\begin{aligned} x' &= x + y \tan\theta \\ y' &= y \\ J &= 1 \end{aligned} \tag{3.43}$$

$J = 1$ indicates that the area affected by the transformation is constant.

3.3.2 Resampling

Geometric transformations can be accomplished in two ways:

1. The actual locations of the picture elements (pixels) are changed, but the elements retain their intensity values. Because of its limited accuracy, this method is only used for simple geometric corrections (aspect ratio and skew).
2. The image is resampled. A digital image defined on an equally spaced grid is converted to a picture on a transformed equally spaced grid. A feature at grid point (x', y') in the geometrically transformed image is located at point (x, y) in the distorted image. In general, the location (x', y') will not coincide with grid points in

the input image. The intensity values of the pixels on the output grid must be determined by interpolation using neighboring pixels on the input grid.

The basic approach is to model the original image defined on the input grid and then to resample this modeled scene, to yield an image with the desired geometric characteristics. For practical reasons the scene is only modeled locally, i.e., in the neighborhood of the interpolation site. The resampling affects image quality (loss of spatial resolution and photometric accuracy).

A digital image, represented by a digital picture function $g_s(j, k)$, is defined on an equally spaced grid $(j\Delta x, k\Delta y)$, $j=1, \ldots, M$, $k=1, \ldots, N$. For an output grid point (x, y), where

$$\left.\begin{aligned} j\Delta x \leq x < (j+1)\Delta x \\ k\Delta y \leq y < (k+1)\Delta y \end{aligned}\right\} \tag{3.44}$$

and if g_s is band-limited by U and V, $g(x, y)$ can be reconstructed exactly by applying the two-dimensional sampling theorem. (See sec. 2.7.)

$$g(x, y) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} g_s(j\Delta x, k\Delta y)\, h(x-j\Delta x, y-k\Delta y) \tag{2.147}$$

If the sampling intervals are chosen to be $\Delta x = 1/(2U)$, $\Delta y = 1/(2V)$, the reconstruction filter is given by

$$h(x, y) = UV \frac{\sin 2\pi Ux}{2\pi Ux} \frac{\sin 2\pi Vy}{2\pi Vy} \tag{2.150}$$

If $\Delta x < 1/2U$ and $\Delta y < 1/2V$, other functions h can be used to represent g exactly through equation (2.147). Equations (2.147) and (2.150) represent the Nyquist–Shannon expansion for band-limited functions. To implement this interpolation formula on a computer, the sum has to be made finite, or equivalently, $h(x, y) = 0$ outside an interval that must include the origin. The right-hand side of equation (2.147) then does not represent $g(x, y)$ exactly. Let $I_n(x, y)$ be such an approximation of $g(x, y)$, given by

$$I_n(x, y) = \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} g_s(j\Delta x, k\Delta y)\, h_n(x-j\Delta x, y-k\Delta y) \tag{3.45}$$

with $h_n(x, y) = 0$ for $|x| > \gamma$, $|y| > \delta$. Depending on the choice of $h_n(x, y)$, various interpolation schemes can be implemented, which differ in accuracy and speed.

1. *Nearest-Neighbor Interpolation* $(n=1)$—In this first approximation, the value of the nearest pixel to (x, y) in the input grid is assigned

to $I_1(x, y)$. Figure 3.14 shows the function h_1 for one dimension. If j=integer $(x+0.5)$ and k=integer $(y+0.5)$, then, for $\Delta x=\Delta y=1$

$$I_1(x, y)=g_s(j, k) \tag{3.46}$$

(j=integer (x) means j is the largest integer number not greater than x). The resulting intensity values correspond to true input pixel values, but the geometric location of a pixel may be inaccurate by as much as $\pm 1/2$ pixel spacings. The sudden shift of true pixel values causes a blocky appearance of linear features. Nearest-neighbor interpolation is used to correct for scanner line length variations in Landsats 1 and 2 digital MSS images by inserting or deleting pixels at appropriate intervals (synthetic pixels [19]). The synthetic pixels may cause misregistration when comparing two MSS images of the same scene taken at different times and should, therefore, be removed by preprocessing. The computational requirements of nearest-neighbor interpolation are relatively low, because only one data value is required to determine a resampled pixel value.

2. *Bilinear Interpolation* $(n=2)$—Bilinear interpolation involves finding the four pixels on the input grid closest to (x, y) on the output grid and obtaining the value of $I_2(x, y)$ by linear approximation, i.e., by assuming that the picture function is linear in the interval $[(j\Delta x, (j+1)\Delta x), (k\Delta y, (k+1)\Delta y)]$. Figure 3.15 shows the function h_2 for one dimension. The approximated value is given by

$$\begin{aligned} I_2(x, y)=&(1-\alpha)(1-\beta)g_s(j, k)+\alpha(1-\beta)g_s(j+1, k)\\ &+\beta(1-\alpha)g_s(j, k+1)+\alpha\beta\, g_s(j+1, k+1) \end{aligned} \tag{3.47}$$

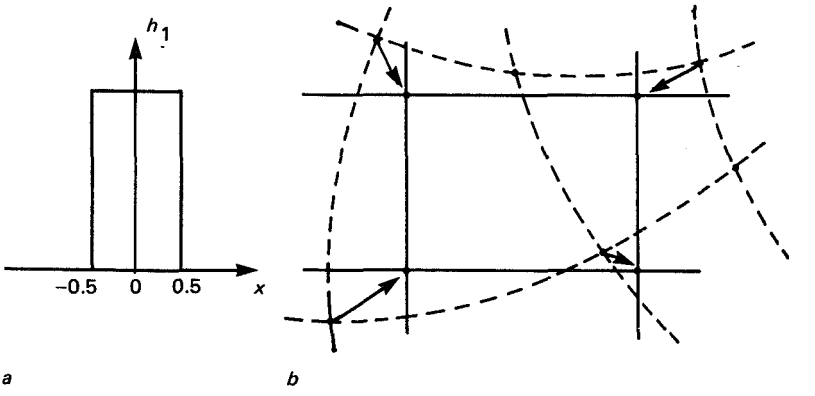


FIGURE 3.14—*Nearest-neighbor interpolation. (a) Interpolation function. (b) Resampled image grid.*

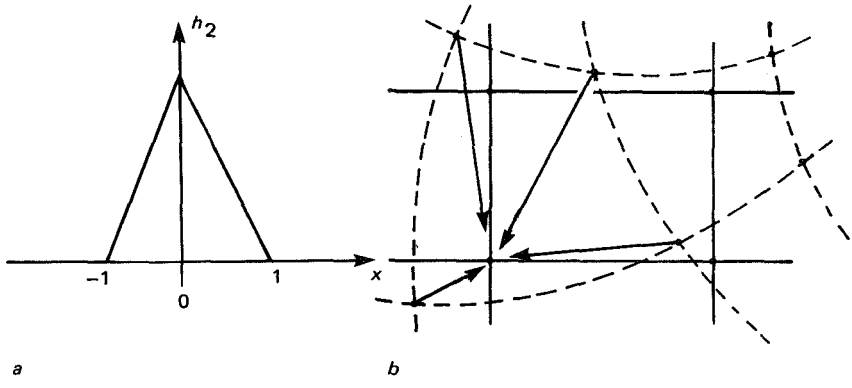


FIGURE 3.15—*Linear interpolation. (a) Interpolation function. (b) Resampled image grid.*

where

$$\left.\begin{aligned} j&=\text{integer } (x) \qquad \alpha=x-j \\ k&=\text{integer } (y) \qquad \beta=y-k \end{aligned}\right\} \tag{3.48}$$

Bilinear interpolation may cause a small loss of image resolution due to the smoothing or blurring nature of linear interpolation. On the other hand, the blocky appearance of linear features, associated with nearest-neighbor interpolation, is reduced. The computational cost of this method is higher because of the additional operations required.

3. *Bicubic Interpolation*—Resampling accuracy can be further improved by modeling the picture locally by a polynomial surface. Use of this polynomial surface implies an increase of the domain of the sampling function $h_n(x, y)$. Frequently a cubic approximation of the ideal sampling function given in equation (2.150) with the 16 nearest neighbors of (x, y) is employed [18]. Such a cubic approximation, which is continuous in value and slope, is given in one dimension (fig. 3.16) by

$$h_3(x)=\begin{cases} 1-2|x|^2+|x|^3 & 0\leq|x|<1 \\ 4-8|x|+5|x|^2-|x|^3 & 1\leq|x|<2 \\ 0 & x|\geq 2 \end{cases} \tag{3.49}$$

A possible implementation is first to interpolate in the line direction to obtain pixel values at locations $(x, k-1)$, (x, k), $(x, k+1)$, $(x, k+2)$ and then to interpolate in the sample direction to obtain $I_3(x, y)$. Using the interpolation function in equation (3.49) for two dimensions and the appropriate pixel values on the input grid yields

$$\begin{aligned} I(x, m)=&-\alpha(1-\alpha)^2\, g_s(j-1, m)+(1-2\alpha^2+\alpha^3)\, g_s(j, m) \\ &+\alpha(1+\alpha-\alpha^2)\, g_s(j+1, m)-\alpha^2(1-\alpha)\, g_s(j+2, m) \\ &m=k-1, k, k+1, k+2 \end{aligned} \tag{3.50}$$

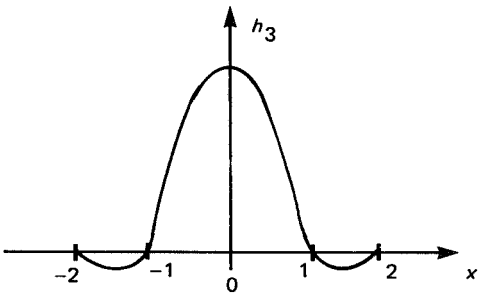


FIGURE 3.16—*Cubic interpolation.*

The final interpolated value $I_3(x, y)$ is obtained by

$$I_3(x,y) = -\beta(1-\beta)^2 I(x,k-1) + (1-2\beta^2+\beta^3) I(x,k) + \beta(1+\beta-\beta^2) I(x,k+1) + \beta^2(\beta-1) I(x,k+2) \quad (3.51)$$

where α, β and j, k are defined as in equation (3.48). Bicubic interpolation is free of the dislocaton of values characteristic of nearest-neighbor interpolation, and the resolution degradation associated with bilinear interpolation is reduced. An important application of bicubic interpolation is the generation of magnified or zoomed image displays. The replication of pixels with nearest-neighbor interpolation is distracting to the eye, but bicubic interpolation represents fine image detail much better. Although higher order interpolation improves the visual appearance of images, analysis techniques such as classification (see ch. 8) may be sensitive to the interpolation method employed.

Resampling results in an image degradation due to the attenuation of higher spatial frequencies caused by the interpolation function and by the aliasing effects associated with discrete interpolation. Interpolation is fundamentally a low-pass filtering operation [20]. The attenuation of higher spatial frequencies, which causes blurring, is a function of the distance of the resampled pixels from the original sampling sites. If the resampling and sampling sites coincide, there is no amplitude attenuation. At a distance of 0.5 pixel, the attenuation of high spatial frequencies is significant. Figure 3.17 shows the modulation transfer functions corresponding to h_n for nearest-neighbor, bilinear, and bicubic interpolation; H_4 corresponds to a truncated form of the ideal $(\sin x)/x$ interpolation function. Aliasing effects (see sec. 2.6.1.1) result from the extension of the interpolator transfer function beyond the cutoff frequency (see fig. 3.17) and cause ringing near edges in the resampled image.

The effects of resampling in the course of geometric corrections are illustrated in figure 3.18, which shows a Landsat MSS false-color image corrected for skew, scale, and rotational errors. The nearest-neighbor interpolation retains the original intensity values, but straight features

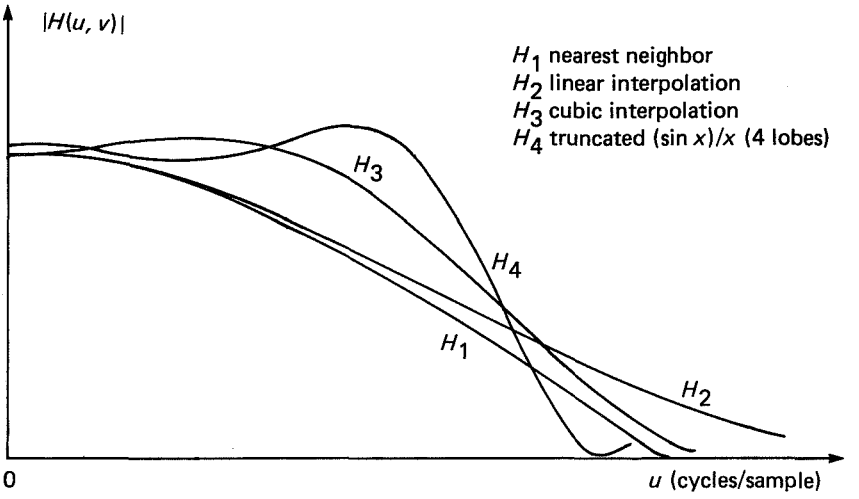


FIGURE 3.17—*Image degradation caused by resampling when sampling site is midway between original samples.*

(airport runway, roads) appear stepped (fig. 3.18*a*). Bilinear interpolation represents some linear features much better (fig. 3.18*b*). Again the airport runways appear jagged, but the white line indicating freeway construction northeast from the river in the upper left section of the image is well represented. With bicubic interpolation, straight lines appear much smoother and edges are sharper (fig. 3.18*c*).

3.4 Radiometric Restoration

The preprocessed and geometrically corrected image $g_R(x', y')$ is related to the ideal object distribution $f(x', y')$ by

$$g_R(x', y') = T_R f(x', y') = h(x', y') * f(x', y') + n_r(x', y') \quad (2.80)$$

where the operator T_R represents the radiometric degradation. The degradation function $h(x-\xi, y-\eta)$ is the PSF of the linearized space-invariant imaging system, and $n_r(x, y)$ represents the random noise in the recorded image. The assumptions of linear image formation and sensor response, and of additive signal-independent noise permit a simpler mathematical treatment of radiometric restoration. (See sec. 2.3.)

The problem of radiometric restoration is the determination of an estimate $\hat{f}$ of the original object distribution f, given the preprocessed and geometrically corrected image g_R and knowledge about the PSF h [21]. Mathematically, the problem is to find the inverse operator T_R^{-1} such that

$$\hat{f} = T_R^{-1} g_R \quad (3.52)$$

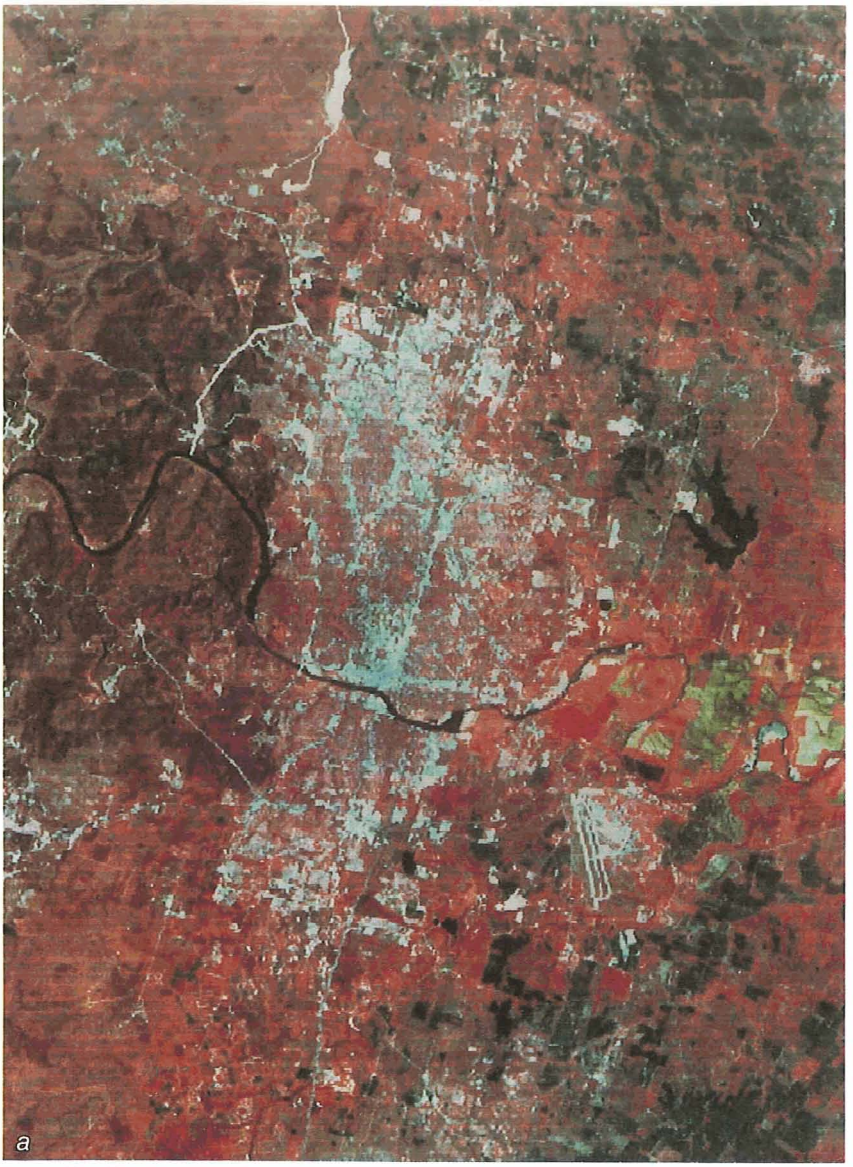


FIGURE 3.18a—*Geometrically corrected Landsat MSS image (scene 1974–16130) resampled with nearest-neighbor interpolation.*

FIGURE 3.18*b*—*Image in part a resampled with bilinear interpolation.*

FIGURE 3.18c—*Image in part a resampled with bicubic interpolation.*

Even if the inverse operator T_R^{-1} exists and is unique, the radiometric image restoration problem is ill conditioned, which means that small perturbations in g_R can produce nontrivial perturbations in $\hat{f}$ [21–23]. Thus, inherent data perturbations in the recorded image can cause undesirable effects in the image restored by inverse transformation. Because random noise is always present, there is an irrecoverable uncertainty in the restored object distribution f.

Discretization of the linear model, equation (2.80), for digital processing leads to the matrix vector equation (see sec. 2.6.2):

$$\mathbf{g}=\mathbf{B}\,\mathbf{f}+\mathbf{n} \tag{2.140}$$

where $\mathbf{g}$, $\mathbf{f}$, and $\mathbf{n}$ are vectors created from the sampled image, sampled object, and sampled noise fields; and $\mathbf{B}$ is a matrix resulting from the sampled PSF h. Because of the nature of the mathematical problem, the matrix $\mathbf{B}$ is always ill conditioned (nearly singular). The solution of the digital radiometric restoration problem is thus tied to the solution of ill-conditioned systems of linear equations. In the presence of noise, the matrix $\mathbf{B}$ can become singular within the bounds of uncertainty imposed by the noise.

Both deterministic and stochastic approaches can be taken to solve the radiometric restoration problem. The deterministic approach implies the solution of a system of linear equations (2.140), but the stochastic approach implies the estimation of a vector subject to random disturbances.

Because of noise and ill conditioning, there is no unique solution to equation (2.140), and some criterion must be used to select a specific solution from the infinite family of possible solutions. For the deterministic approach, radiometric restoration can be posed as an optimization problem. For example, a possible criterion for solution is minimization of the noise $\mathbf{n}$. This approach leads to a least-squares problem, where it is necessary to find a solution $\hat{\mathbf{f}}$ such that

$$\mathbf{n}^T\mathbf{n}=(\mathbf{g}-\mathbf{B}\,\hat{\mathbf{f}})^T(\mathbf{g}-\mathbf{B}\,\mathbf{f}) \tag{3.53}$$

is minimized. In the stochastic approach to restoration it is assumed that the object distribution $\mathbf{f}$ is a random field. A criterion of solution is to construct an estimate such that over the ensemble of estimates and object distributions the expected value of the difference between estimate and original object distribution is minimized. This approach is equivalent to finding a solution $\hat{\mathbf{f}}$ such that the error ϵ is a minimum, where

$$\epsilon=E\{(\mathbf{f}-\hat{\mathbf{f}})^T(\mathbf{f}-\hat{\mathbf{f}})\} \tag{3.54}$$

These two criteria lead to the inverse filter and the Wiener filter, respectively.

It was shown in section 2.5 that for digitized images the fundamental limit on the size of image detail is determined by the Nyquist frequency

in equation (2.104). Radiometric restoration can recover detail below the Nyquist frequency. Because of the presence of noise, radiometric restoration has to consider the tradeoff between sharpness of the restored image and the amount of noise in it. In addition, the criterion should insure positive restoration because the original image is everywhere positive. The restoration filters to be discussed may create negative intensity values in the restored image, values which have no physical meaning. An excellent survey of positive restoration methods, which now require an excessive amount of computation time, was prepared by Andrews [24].

Mathematically, equation (2.80) is a Fredholm integral equation of the first kind, which tends to have an infinite number of solutions. To obtain a unique solution, some constraints must be imposed. A method described in [22] constrains the mean square error to a certain value and determines the smoothest solution with that error. Another technique [25] uses the constraint that the restored image is everywhere positive.

3.4.1 Determination of Imaging System Characteristics

Solution of the radiometric restoration problem requires knowledge of the PSF $h(x, y)$ or the corresponding transfer function $H(u, v)$. Determination of the transfer characteristics is a classical problem in dynamic systems analysis. If the imaging system is available, the transfer characteristics may be determined by measuring the response of the system to specific test patterns. An example is the work carried out at the Jet Propulsion Laboratory, where extensive measurements of vidicon camera systems were made before their launch [26]. For space-invariant systems, it is sometimes possible to postulate a model for the camera system and to calculate the PSF $h(x, y)$ or transfer function $H(u, v)$. The PSFs for some typical degradations were derived under simplifying assumptions in sec. 2.4.

If the imaging system is too complex for analytic determination of $h(x, y)$, or if the degrading system is unavailable, $h(x, y)$ must be estimated from the degraded picture itself. The existence of sharp points in the original scene can be used to measure the PSF directly in the image. (The PSF is the response of a linear imaging system to a point source.) For example, in astronomical pictures, the image of a point star could be used as an estimate of the PSF.

Usually, natural scenes will not contain any sharp points, but they may contain edges in various orientations. In this case, the PSF can be determined from the derivatives of the images of these edges. An edge is an abrupt change in brightness. The response to such a step function in any direction is called the edge-spread function (ESF) $h_e(x, y)$. The derivative of the ESF in any direction is called the line-spread function (LSF) $h_l(x, y)$ in that direction. The LSF in any direction is the integral of the PSF in that direction. Thus, the derivatives of edge images in different

orientations can be used to reconstruct the PSF of the imaging system. The problem with this method is that noise strongly effects the values of the derivative [27–29].

3.4.2 Inverse Filter

Minimization of equation (3.53), yields the least-squares estimate

$$\hat{\mathbf{f}}=\mathbf{B}^{-1}\mathbf{g} \tag{3.55}$$

which is the inverse filter restoration. The numerical solution of equation (3.55) is in general impossible. Suppose that the image was digitized with $N=512$. Inversion of a 262,144 by 262,144 matrix is required. However, for a space-invariant imaging system, the matrix $\mathbf{B}$ is block circulant. A matrix with this structure is diagonalized by the two-dimensional DFT, leading to

$$\hat{f}=\mathcal{F}^{-1}\left\{\frac{G}{H}\right\} \tag{3.56}$$

$\mathcal{F}$ is the discrete two-dimensional Fourier transform operator, $\hat{f}$ is the restored image matrix; and G and H are the Fourier transforms of g and h, respectively. The same result can be obtained by direct application of the convolution property of the Fourier transform to equation (2.80) if $n_r=0$.

The ill-conditioned nature of image restoration is preserved by the use of the least-squares criterion. The elements of H are an approximation to the eigenvalues of $\mathbf{B}$. As the eigenvalues become small (near singularity), the inverse becomes large, and an amplification of noise results. For high signal-to-noise ratios and with small amounts of image blur, the inverse filter performs well, provided there are no zeros in H. Because the convolution performed by the DFT is circular, there is an edge effect from convolution wraparound at the image borders. The addition of zeros before the transformation of f and h results in a suppression of wraparound. (See sec. 2.6.2.)

Because of its computational simplicity the inverse filter

$$H_I(u, v)=\frac{1}{H(u, v)} \tag{3.57}$$

is used for radiometric restoration of remotely sensed images where atmospheric turbulence is the main cause of degradation. If, however, the transfer function H has zeros for defocus degradation (see sec. 2.4.3), the inverse filter approaches infinity. Even for transfer functions without any zeros, the inverse filter enhances noise to an extent that will interfere with visual interpretation more than does a loss of fine image detail.

Therefore, a modified inverse filter, shown in figure 3.19, is often used [10], where

$$H_I(u, v) = \begin{cases} \dfrac{1}{H(u, v)} & u, v \leq S \\ H_M & u, v > S \end{cases} \tag{3.58}$$

The limitations depend on the signal-to-noise ratio of the image to be processed and is determined empirically.

3.4.3 Optimal Filter

The optimal or Wiener restoration filter results from the minimization of equation (3.54) in the stochastic approach. The problem becomes mathematically tractable if a linear estimate is chosen, that is, if a linear relationship between the estimate $\hat{\mathbf{f}}$ and the gray values in $\mathbf{g}$ is assumed. The problem is then to find a restoration filter $\mathbf{W}$ such that the estimate $\hat{\mathbf{f}}$ minimizes the error measure in equation (3.54), where

$$\hat{\mathbf{f}} = \mathbf{W}\mathbf{g} \tag{3.59}$$

Because signal-independent noise was assumed,

$$E\{\mathbf{f}\mathbf{n}^T\} = E\{\mathbf{n}^T\mathbf{f}\} = 0 \tag{3.60}$$

and the mean-square-error solution is [30–32]

$$\mathbf{W} = \mathbf{R}_{ff}\mathbf{B}^T(\mathbf{B}\mathbf{R}_{ff}\mathbf{B}^T + \mathbf{R}_{nn})^{-1} \tag{3.61}$$

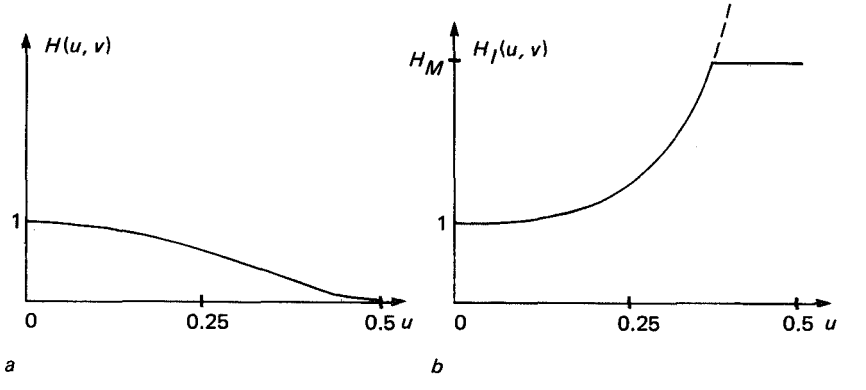


FIGURE 3.19—*Inverse filter. (a) Modulation transfer function. (b) Modified inverse filter.*

where

$$\left.\begin{aligned}\mathbf{R}_{ff}&=E\{\mathbf{f}\mathbf{f}^T\}\\ \mathbf{R}_{nn}&=E\{\mathbf{n}\mathbf{n}^T\}\end{aligned}\right\} \tag{3.62}$$

are the autocorrelation matrices of $\mathbf{f}$ and $\mathbf{n}$, respectively; and $\mathbf{R}_{ff}$ and $\mathbf{R}_{nn}$ represent the information about the signal and noise processes, respectively, necessary to carry out the restoration.

The minimum mean-square-error restoration requires very large matrix computations. Conversion to an efficient representation in the spatial frequency domain with the DFT is possible if the underlying random process for $\mathbf{f}$ is stationary, which means that the restored image can be considered a zero-mean process plus an additive constant. With this assumption and for space-invariant imaging systems, all matrices in equation (3.61) can be diagonalized by the DFT. The Fourier transforms of the correlation functions $\mathbf{R}_{ff}$ and $\mathbf{R}_{nn}$ are the spectral densities S_{ff} and S_{nn}, respectively, in equation (2.18). The optimal or Wiener filter in the frequency domain is obtained from equation (3.61) as

$$W(u, v)=\frac{H(u, v)^*}{|H(u, v)|^2+\dfrac{S_{nn}(u, v)}{S_{ff}(u, v)}} \tag{3.63}$$

where H^* is the complex conjugate of H. The estimate $\hat{f}(x, y)$ of the restored image is thus

$$\hat{f}(x, y)=\mathcal{F}^{-1}\left\{\frac{H(u, v)^*G(u, v)}{|H(u, v)|^2+\dfrac{S_{nn}(u, v)}{S_{ff}(u, v)}}\right\} \tag{3.64}$$

There is no ill-conditioned behavior associated with the optimal filter. Even though $H(u, v)$ may have zero elements, the denominator in equation (3.63) is then determined by the ratio S_{nn}/S_{ff}. Thus, a restored image can be generated if the matrix $\mathbf{B}$ is singular and noise is present. In fact, the presence of noise makes the restoration possible. If noise approaches zero ($S_{nn} \rightarrow 0$), the optimal filter becomes the inverse filter. The visual appearance of optimally restored images for low signal-to-noise ratios is often not good. This deficiency may be due to using a linear estimate and a mean-square-error criterion. The error criterion should also take into account the frequency response of the human visual system and its logarithmic response to varying light intensities, and it should insure positive restoration.

3.4.4 Other Radiometric Restoration Techniques

Radiometric restoration without knowledge of the PSF may be achieved by homomorphic filtering [33, 34]. Here the recorded image is mapped

from its original representation into another domain where the original image and the degrading function are additively related. Such a transformation is the Fourier transform, which maps convolution into multiplication, followed by the complex logarithm, which maps multiplication into addition [35, 36]. The inverse transform is the complex exponential function followed by the inverse Fourier transform. The restoration criterion is to find a linear operator W such that the power spectral densities of the restored and of the original image are equal. In the spatial frequency domain

$$\hat{F}(u, v) = G(u, v) W(u, v) \tag{3.65}$$

with the criterion

$$S_{\hat{f}\hat{f}}(u, v) = S_{ff}(u, v) \tag{3.66}$$

The homomorphic restoration filter is

$$\begin{aligned} W(u, v) &= \left(\frac{S_{\hat{f}\hat{f}}(u, v)}{S_{gg}(u, v)} \right)^{1/2} \\ &= \left(\frac{S_{ff}(u, v)}{|H(u, v)|^2 S_{ff}(u, v) + S_{nn}(u, v)} \right)^{1/2} \end{aligned} \tag{3.67}$$

which is obtained without detailed knowledge of H and S_{nn}. Hunt [37] proposed a constrained least-squares filter that does not require knowledge of the spectral densities. The restoration filter is

$$W(u, v) = \frac{H^*(u, v)}{|H(u, v)|^2 + \gamma |C(u, v)|^2} \tag{3.68}$$

where C is a constraint matrix. The parameter γ is determined by iteration.

The Wiener filter requires a maximum amount of a priori information, namely, the spectral densities of f and n. The constrained least-squares filter does not require knowledge of the power spectra. In the homomorphic filtering approach the PSF is estimated from the degraded image. The inverse filter requires no a priori information, and this fact is reflected in the usually poor quality of the restoration.

REFERENCES

[1] Turner, R. E.; et al.: Influence of the Atmosphere on Remotely Sensed Data. Proceedings of Conference on Scanners and Imagery Systems for Earth Observations, SPIE J., vol. 51, 1974, pp. 101–114.

[2] Chavez, P.: Atmospheric, Solar, and MTF Corrections for ERTS Digital Imagery, Proc. Am. Soc. Photogrammetry, Oct. 1975, pp. 69–69a.

[3] Rogers, R. H.; and Peacock, K.: A Technique for Correcting ERTS Data for Solar and Atmospheric Effects. Symposium on Significant Results Obtained from the Earth Resources Technology Satellite-1, NASA SP-327, Washington, D.C., 1973, pp. 1115–1122.

[4] Murray, W. L.; and Jurica, J. G.: The Atmospheric Effect in Remote Sensing of Earth Surface Reflectivities, Laboratory for Applications of Remote Sensing Information. Note 110273, Purdue University, Lafayette, Ind., 1973.
[5] Fraser, R. S.: Computed Atmospheric Corrections for Satellite Data. Proceedings of Conference on Scanners and Imagery Systems for Earth Observations, SPIE J., vol. 51, 1974, pp. 64–72.
[6] Potter, J.; and Sheldon, M.: Effect of Atmospheric Haze and Sun Angle on Automatic Classification of ERTS-1 Data. Proceedings of the Ninth International Symposuim on Remote Sensing of Environment, 1974.
[7] Hammond, H. K.; and Mason, H. L.: Precision Measurement and Calibration. NBS Special Publication 300, vol. 7, 1971 (Order No. C13.10:300/V.7).
[8] Advanced Scanners and Imaging Systems for Earth Observations. NASA SP–335, Washington, D.C., Dec. 1972.
[9] Papoulis, A.: The Fourier Integral and Its Applications. McGraw-Hill, New York, 1962.
[10] Seidman, J.: Some Practical Applications of Digital Filtering in Image Processing. Proceedings of Computer Image Processing and Recognition, University of Missouri, Columbia, Mo., Aug. 1972.
[11] Rindfleisch, T. C., et al.: Digital Processing of the Mariner 6 and 7 Pictures, J. Geophys. Res., vol. 76, 1971, pp. 394–417.
[12] Goetz, A. F. H.; Billingsley, F. C.; Gillespie, A. R.; Abrams, M. J.; and Squires, R. L.: Application of ERTS Images and Image Processing to Regional Geologic Problems and Geologic Mapping in Northern Arizona. NASA/JPL TR 32–1597, May 1975.
[13] Horn, B. K. P.; and Woodham, R. J.: Destriping Satellite Images. Artificial Intelligence Lab. Rep. AI 467, Massachusetts Institute of Technology, Cambridge, Mass., 1978.
[14] Puccinelli, E. F.: Ground Location of Satellite Scanner Data, Photogr. Eng. and Remote Sensing, vol. 42, 1976, pp. 537–543.
[15] Mottershead, C. T.; and Phillips, D. R.: Image Navigation for Geosynchronous Meteorological Satellites. Seventh Conference on Aerospace and Aeronautical Meteorology and Symposium on Remote Sensing from Satellites, American Meteorological Society, Melbourne, Fla. 1976, pp. 260–264.
[16] ERTS Data Users Handbook. Doc. 71SD4249, NASA. Washington, D.C., 1972. Appendix B.
[17] Caron, R. H.; and Simon, K. W.: Attitude Time Series Estimator for Rectification of Spaceborn Imagery, J. Spacecr., Rockets, vol. 12, 1975, pp. 27–32.
[18] Rifman, S. S.: Digital Rectification of ERTS Multispectral Imagery. Symposium on Significant Results Obtained from the Earth Resources Technology Satellite-1, NASA SP–327, Washington, D.C., 1973, pp. 1131–1142.
[19] Thomas, V. L.: Generation and Physical Characteristics of the Landsat 1 and 2 MSS Computer Compatible Tapes. NASA/GSFC Report X-563-75-223, Nov. 1975.
[20] Forman, M. L.: Interpolation Algorithms and Image Data Artefacts. NASA/GSFC Report X-933-77-235, Oct. 1977.
[21] Andrews, H. C.; and Hunt, R.: Digital Image Restoration. Prentice Hall, Englewood Cliffs, N. J., 1977.
[22] Twomey, S.: On the Numerical Solution of Fredholm Integral Equations of the First Kind by the Inversion of the Linear System Produced by Quadrature, J. Assoc. Comput. Mach., vol. 10, 1963, pp. 97–101.
[23] Sondhi, M. M.: Image Restoration: The Removal of Spatially Invariant Degradations, Proc. IEEE, vol. 60, 1972, pp. 842–853.
[24] Andrews, H. C.: Positive Digital Image Restoration Techniques—A Survey. Report No. ATR–73(8193)–2, Aerospace Corp., Feb. 1973.
[25] McAdam, D. P.: Digital Image Restoration by Constrained Deconvolution, J. Opt. Soc. Am., vol. 60, 1970, pp. 1617–1627.

[26] O'Handley, D. A.; and Green, W. B.: Recent Developments in Digital Image Processing of the Image Processing Laboratory of the Jet Propulsion Laboratory, Proc. IEEE, vol. 60, 1972, pp. 821–828.
[27] Jones, R. A.; and Yeadon, E. C.: Determination of the Spread Function from Noisy Edge Scans, Photogr. Sci. Eng., vol. 13, 1969, pp. 200–204.
[28] Jones, R. A.: An Automated Technique for Deriving MTF's from Edge Traces, Photogr. Sci. Eng., vol. 11, 1967, pp. 102–106.
[29] Berkovitz, M. A.: Edge Gradient Analysis OTF Accuracy Study, *in* Proceedings of SPIE Seminar on Modulation Transfer Function, Boston, Mass., 1968.
[30] Horner, J. L.: Optical Spatial Filtering with the Least-Mean-Square-Error Filter, J. Opt. Soc. Am., vol. 59, 1969, pp. 553–558.
[31] Helmstrom, C. W.: Image Restoration by the Method of Least Squares, J. Opt. Soc. Am., vol. 57, 1967, pp. 297–303.
[32] Slepian, D.: Linear Least-Squares Filtering of Distorted Images, J. Opt. Soc. Am., vol. 57, 1967, pp. 918–922.
[33] Cole, E. R.: The Removal of Unknown Image Blurs by Homomorphic Filtering. Ph.D. dissertation, Department of Electrical Engineering, University of Utah, Salt Lake City, Utah, June 1973.
[34] Stockham, T. G.: Image Processing in the Context of a Visual Model, Proc. IEEE, vol. 60, 1972, pp. 828–842.
[35] Cannon, T. M.: Digital Image Deblurring by Nonlinear Homomorphic Filtering. Ph.D. thesis, Computer Science Department, University of Utah, Salt Lake City, Utah, Aug. 1974.
[36] Oppenheim, A. V.; Schafer, R. W.; and Stockham, T. G.: Nonlinear Filtering of Multiplied and Convolved Signals, Proc. IEEE, vol. 56, 1968, pp. 1264–1291.
[37] Hunt, B. R.: An Application of Constrained Least Squares Estimation to Image Restoration by Digital Computer, IEEE Trans. Comput., vol. C-22, 1973, pp. 805–812.

Page intentionally left blank

4. Image Enhancement

4.1 Introduction

The goal of image enhancement is to aid the human analyst in the extraction and interpretation of pictorial information. The interpretation is impeded by degradations resulting from the imaging, scanning, transmission, or display processes. Enhancement is achieved by the articulation of features or patterns of interest within an image and by a display that is adapted to the properties of the human visual system. (See sec. 2.8.) Because the human visual system discriminates many more colors than shades of gray, a color display can represent more detailed information than a gray-tone display.

The information of significance to a human observer is definable in terms of the observable parameters contrast, texture, shape, and color [1]. The characteristics of the data and display medium and the properties of the human visual system determine the transformation from the recorded to the enhanced image, and, therefore, the range and distribution of the observable parameters in the resulting image [2–4]. The decisions of which parameter to choose and which features to represent by that parameter are determined by the objectives of the particular application. Enhancement operations are applied without quantitative knowledge of the degrading phenomena, which include contrast attenuation, blurring, and noise. The emphasis is on human interpretation of the pictures for extraction of information that may not have been readily apparent in the original. The techniques try to attenuate or discard irrelevant features and at the same time to emphasize features or patterns of interest [5–7].

Multi-image enhancement operators generate new features by combining components (channels) of multi-images. For multi-images with more than three components, the dimensionality can be reduced to enable an unambiguous color assignment. Enhancement methods may be divided into:

1. Contrast enhancement (gray-scale modification)
2. Edge enhancement
3. Color enhancement (pseudocolor and false color)
4. Multi-image enhancement

Contrast enhancement, edge enhancement, and pseudocolor enhancement are performed on monochrome images or on individual components of multi-images.

4.2 Contrast Enhancement

The goal of contrast enhancement is to produce a picture that optimally uses the dynamic range of a display device. The device may be a television screen, photographic film, or any other equipment used to present an image for visual interpretation. The human eye can simultaneously discriminate only 20 to 30 gray levels [8]. This subjective brightness range does not cover the full gray scale of a display device. For the human eye to see subtle changes in brightness, the contrast characteristics of an image must be adapted to the subjective brightness range. The contrast characteristics of an image are influenced by such factors as camera exposure settings, atmospheric effects, solar lighting effects, and sensor sensitivity. These factors often cause a recorded image not to span the dynamic range to which it is digitized. On the other hand, contrast enhancement is limited by the darkest and the lightest areas in an image.

Gray-scale transformations are applied in a uniform way to the entire picture to stretch the contrast to the full dynamic range. Spatially dependent degradations (e.g., vignetting and shading) may be corrected by spatially nonuniform transformations. Spatially independent gray-scale transformations can be expressed by

$$g_e = T_c g \tag{4.1}$$

where g and g_e are the recorded and the enhanced image with M rows and N columns, respectively, and T_c is a linear or nonlinear gray-scale transformation that is applied to every point in the image separately. The dynamic range of both gray scales is the same; i.e., $0 \leq g(j, k) \leq K$ and $0 \leq g_e(j, k) \leq K$, where $K = 2^b - 1$. The number K is the maximum gray value and b is the number of quantization bits. The quantities $g(j, k)$ and $g_e(j, k)$ are the gray values of g and g_e at row j and column k, respectively.

Piecewise linear transformations may be used to enhance the dark, midrange, or bright region of the gray scale and to correct for display nonlinearities. The range $[l, u]$ in the recorded image may be linearly transformed to the range $[L, U]$ in the enhanced image by

$$g_e(j, k) = \begin{cases} \dfrac{L}{l} g(j, k) & g(j, k) < l \\ \dfrac{U - L}{u - l} [g(j, k) - l] + L & l \leq g(j, k) \leq u \\ \dfrac{K - U}{k - u} [g(j, k) - u] + U & g(j, k) > u \end{cases} \tag{4.2}$$

Figure 4.1*a* shows the gray-scale transformation, represented by equation (4.2).

Sometimes the high and low values of a digital picture represent saturation effects. It may be necessary to find a piecewise linear transformation that causes a specified percentage p_L of the low values and a percentage p_U of the high values to be set to L and U, respectively [1]. In practice L is usually set to zero, and U is set to K. Consider the following constraints:

$$\left.\begin{aligned} \frac{1}{MN}\sum_{z=0}^{l-1} H_g(z) &= p_L \\ \frac{1}{MN}\sum_{z=u+1}^{K} H_g(z) &= p_U \end{aligned}\right\} \tag{4.3}$$

with

$$MN = \sum_{z=0}^{K} H_g(z)$$

They define the gray levels l and u. The function $H_g(z)$ is the frequency of occurrence of gray level z in g and is called the histogram of image g. The transformation is then given by

$$\begin{aligned} g_e(j,k) &= \frac{U-L}{u-l}[g(j,k)-l]+L \qquad && u \le g(j,k) \le l \\ g_e(j,k) &= L && g(j,k) < l \\ g_e(j,k) &= U && g(j,k) > u \end{aligned} \tag{4.4}$$

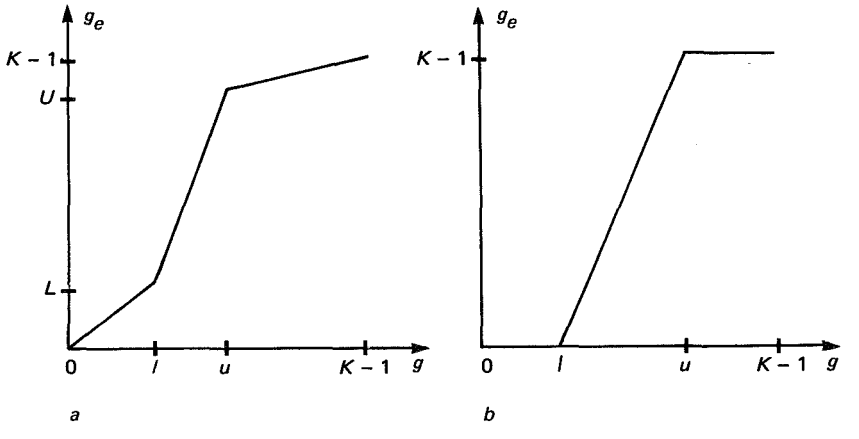


FIGURE 4.1—*Gray-scale transformations for enhancement. (a) Piecewise linear gray-scale transformation. (b) Saturation of low and high values to black and white.*

and shown in figure 4.1*b*. The shape of an image histogram provides information about the contrast characteristics of an image. For example, a narrow histogram indicates a low-contrast image, and a multimodal histogram indicates the existence of regions with different brightness.

Figure 4.2 illustrates linear contrast enhancement. Figure 4.2*a* shows the recorded Landsat MSS 7 image, and figure 4.2*b* is the result of applying the gray-scale transformation, (4.3) and (4.4), with $p_L=1$ percent, and $p_U=1$ percent. The histogram of the recorded image is shown in figure 4.3.

Nonlinear gray-scale transformations may be used to correct for display nonlinearities. Figure 4.4*a* shows a logarithmic transform used to compensate for nonlinear film characteristics. For images with bimodal histograms, each of the histogram zones may be enhanced to the full brightness range by the transformation shown in figure 4.4*b*. The specification of the transformation T_c is facilitated by evaluating the contrast characteristics of a given image from its histogram.

Another important type of contrast enhancement is histogram modification, in which a gray-scale transformation is used to give the picture a specified distribution of gray values [9, 10]. Two frequently used distributions approximate a normally distributed (Gaussian) or a flat (constant) histogram. In pictures with a flat histogram all gray levels occur equally often. Histogram flattening (also called histogram equalization) produces pictures with higher contrast, because the points in the densely populated regions of the gray scale are forced to occupy a larger number of gray levels, resulting in these regions of the gray scale being stretched. Points in sparse regions of the gray scale occupy fewer levels. Figure 4.5 illustrates nonlinear contrast enhancement through histogram modification of the image in figure 4.2*a*. Histogram modification may also be required before comparison of two pictures of the same scene taken at different times. If the pictures were taken under different lighting conditions, the differences can be compensated for by transforming both pictures to a standard histogram.

4.3 Edge Enhancement

Imaging and scanning processes cause blurring degradations. Blurring is an averaging operation, the extent of which is determined by the point spread function (PSF) of the system. Blurred images may therefore be sharpened by a differentiation process. For space-invariant systems the imaging process can be described by the convolution of the original image f with the system PSF h. (See eq. (2.80).) Because of the low-pass characteristics of imaging systems, higher spatial frequencies are weakened more than lower frequencies. Thus, sharpening or edge enhancement can be achieved by high-pass filtering, emphasizing higher spatial frequen-

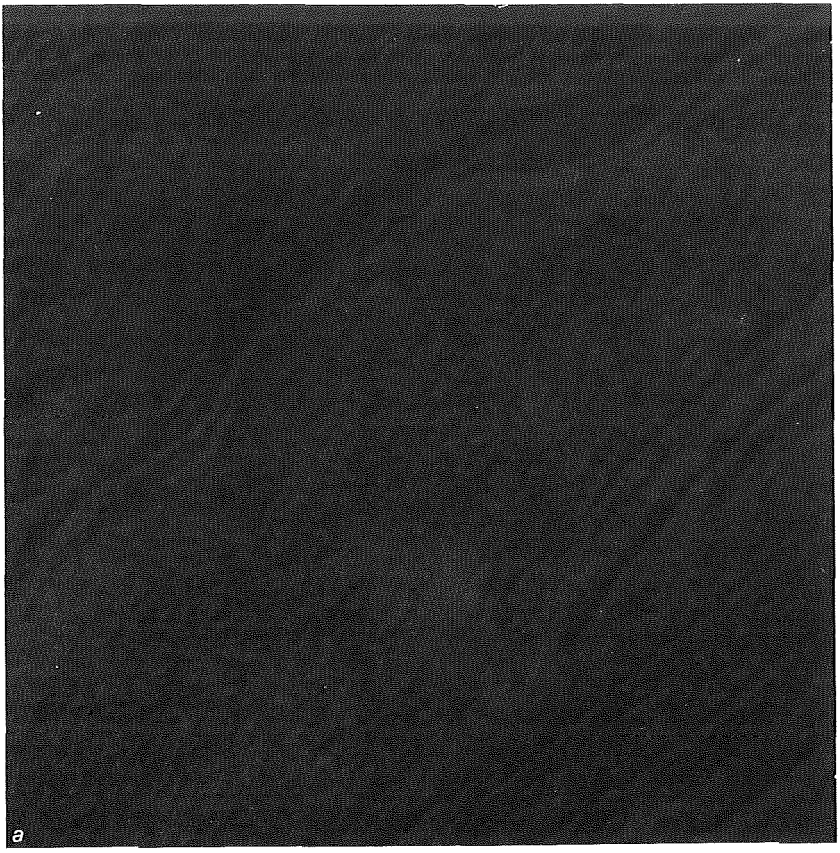


FIGURE 4.2a—*Linear contrast enhancement Recorded Landsat MSS7 image (scene 692-15192).*

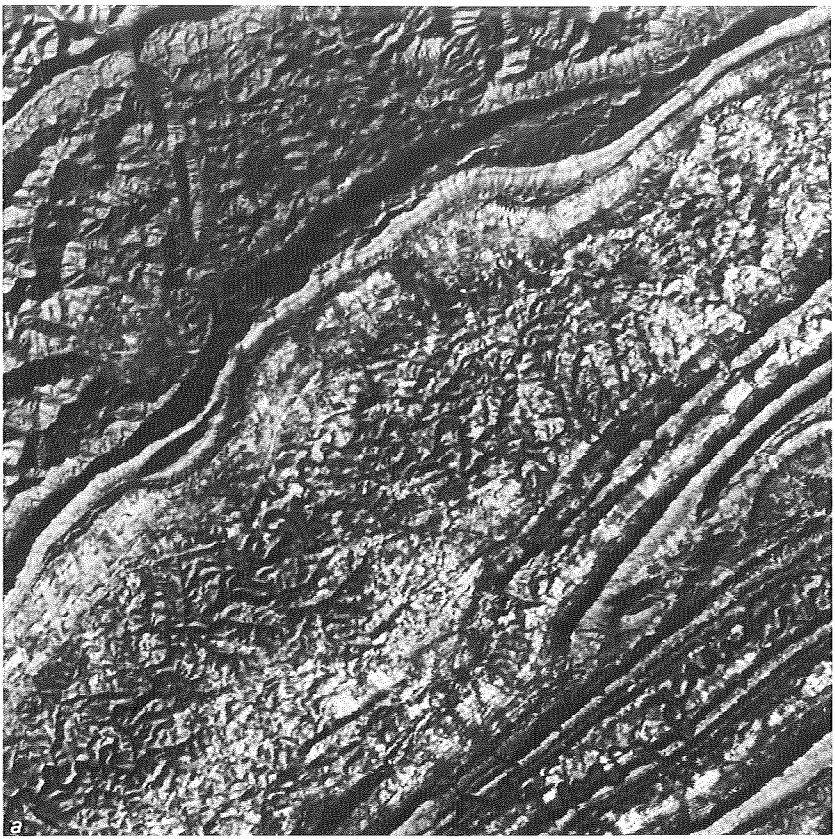


FIGURE 4.2*b*—*Image of part a enhanced with 1 percent of the lowest and highest pixel values set to black and white.*

AVERAGE GRAY LEVEL = 10.476 STANDARD DEVIATION = 4.645 NUMBER OF ELEMENTS = 2391402

FREQUENCY HISTOGRAM NORMALIZED TO MAXIMUM = 100 (LEVELS WITH ZERO FREQUENCY OMITTED)

GRAY	FREQ	PERCENT
0	112332	4.697
1	33	0.001
2	3414	0.143
3	46977	1.964
4	83584	3.492
5	99155	4.146
6	119256	4.987
7	149976	6.272
8	159485	6.669
9	173631	7.261
10	192254	8.039
11	217842	9.109
12	214855	8.984
13	189253	7.914
14	172621	7.218
15	151938	6.354
16	94739	3.962
17	64699	2.705
18	55599	2.325
19	42260	1.767
20	25267	1.057
21	12025	0.503
22	4853	0.203
23	2614	0.109
24	1366	0.057
25	590	0.025
26	268	0.011
27	137	0.006
28	62	0.003
29	41	0.002
30	39	0.002
31	32	0.001
32	38	0.002
33	33	0.001
34	22	0.001
35	29	0.001
36	26	0.001
37	16	0.001
38	21	0.001
39	14	0.001
40	8	0.000
41	14	0.001
42	14	0.001
43	7	0.000
44	3	0.000
45	8	0.000
46	8	0.000
47	5	0.000
48	6	0.000
49	3	0.000
50	2	0.000
51	3	0.000
52	1	0.000
53	1	0.000
55	1	0.000
57	1	0.000

FIGURE 4.3—*Histogram of image in figure 4.2a.*

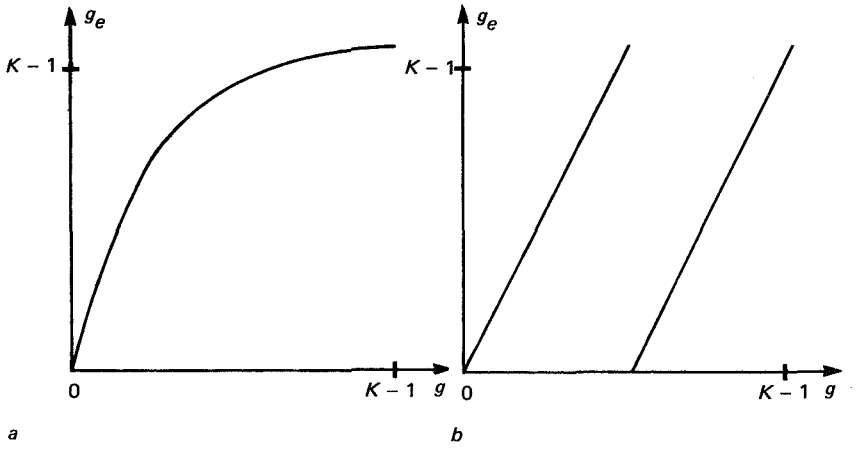


FIGURE 4.4—*Nonlinear contrast enhancement. (a) Logarithmic mapping to correct for film nonlinearity. (b) Enhancement of different intensity ranges to full brightness.*

cies without quantitative knowledge of the PSF [9]. Filtering may be performed in the spatial domain or by multiplication of Fourier transforms in the frequency domain. (See sec. 2.2.5.)

When a picture is blurred and noisy, differentiation or high-pass filtering cannot be used indiscriminately for edge enhancement. Noise generally involves high rates of change of gray levels and hence high spatial frequencies. Sharpening enhances the noise. Therefore, the noise should be reduced or removed before edge enhancement.

Simple differentiation operators are the gradient and the Laplacian [9]. The magnitude of the digital gradient at line j and column k of an image g is defined by

$$|\nabla g(j,k)| = \sqrt{[\Delta_x g(j,k)]^2 + [\Delta_y g(j,k)]^2} \tag{4.5}$$

and its direction is $\phi(j,k)$, where

$$\phi(j,k) = \tan^{-1}\frac{\Delta_y g(j,k)}{\Delta_x g(j,k)} \tag{4.6}$$

where

$$\begin{aligned} \Delta_x g(j,k) &= g(j,k) - g(j-1,k) \\ \Delta_y g(j,k) &= g(j,k) - g(j,k-1) \end{aligned} \tag{4.7}$$

The quantities $\Delta_x g(j,k)$ and $\Delta_y g(j,k)$ are the first differences in the row and column directions, respectively. An edge-enhanced image g_e is obtained by

$$g_e = |\nabla g| \tag{4.8}$$

FIGURE 4.5a—*Flat histogram of image in figure 4.2a.*

FIGURE 4.5*b*—*Normally distributed histogram of image in figure 4.2a.*

Figure 4.6 shows the results of applying the digital gradient operator to the image in figure 4.2*a*. Each element in figure 4.6*a* represents the magnitude of the gradient, given by equation (4.5), at this pixel location, and each element in figure 4.6*b* represents the direction of the gradient, given by equation (4.6). Black is the direction to the neighbor on the left, and gray values of increasing lightness represent directions of increasing counterclockwise orientation. The digital Laplacian at location (j, k) of an image g is given by

$$\nabla^2 g(j, k) = g(j+1, k) + g(j-1, k) + g(j, k+1) + g(j, k-1) - 4g(j, k) \tag{4.9}$$

An edge enhanced g_e is obtained as

$$g_e = \nabla^2 g \tag{4.10}$$

Another method of edge enhancement is to subtract the Laplacian from the blurred image g, yielding the enhanced image as [9]:

$$g_e = g - \nabla^2 g \tag{4.11}$$

An element of g_e is given by

$$g_e(j, k) = 5g(j, k) - [g(j+1, k) + g(j-1, k) + g(j, k+1) + g(j, k-1)] \tag{4.12}$$

Edge enhancement may also be performed by filtering. (See sec. 2.2.5.) The enhanced image is obtained by convolving g with a filter function h

$$g_e = g * h \tag{4.13}$$

Filtering in the frequency domain is performed by multiplication of the Fourier transforms of g and h, G and H respectively, and inverse transformation of the product:

$$g_e = \mathcal{F}^{-1}\{GH\} \tag{4.14}$$

The design of enhancement filters can be performed intuitively in the spatial frequency domain because of the relationship between sharpness and spatial frequencies. (See sec. 2.2.3.1.) Frequency-domain filtering also permits the enhancement of features in a specific direction. This enhancement is possible because the two-dimensional Fourier transform contains information about the direction of features. High spatial frequencies in a certain direction in the frequency spectrum indicate sharp features orthogonal to that direction in the original image.

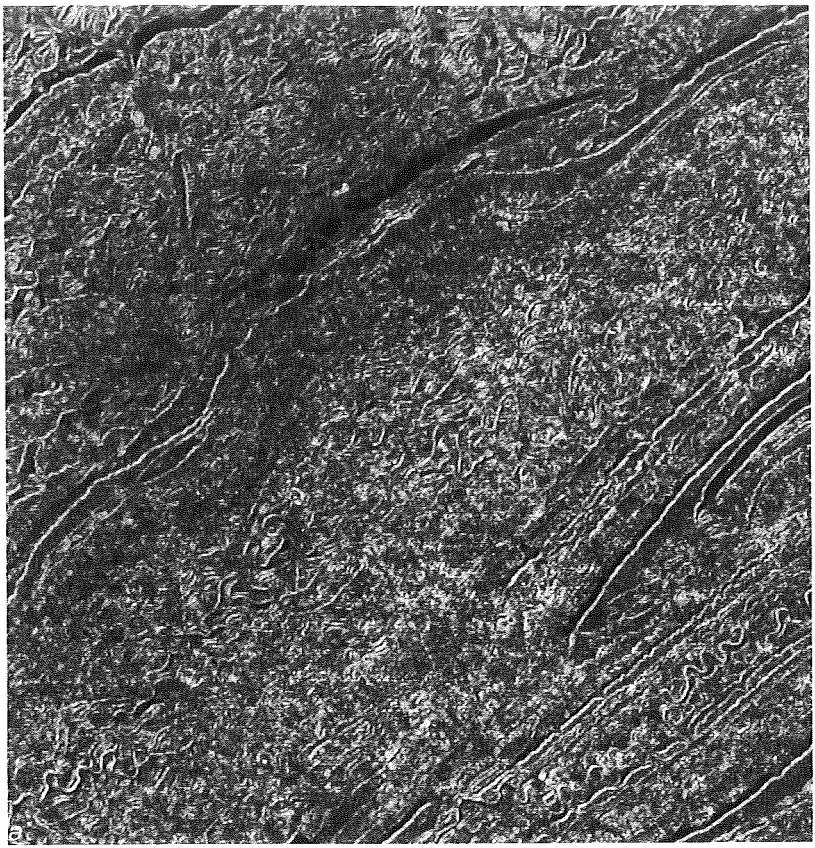


FIGURE 4.6a—*Edge enhancement by differentiation. Each element represents the magnitude of the gradient.*

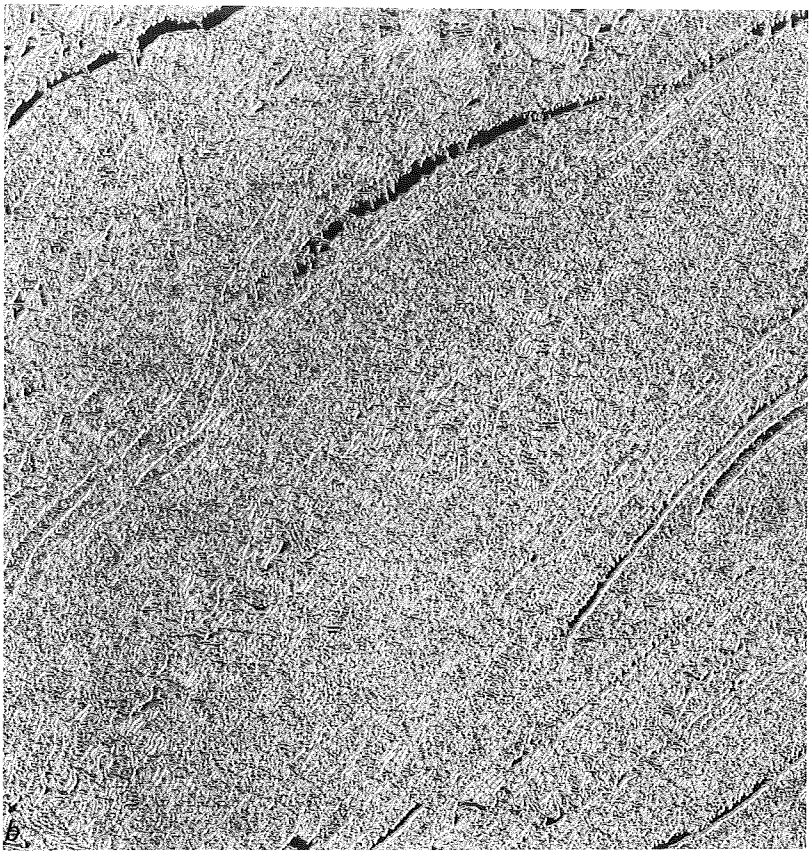


FIGURE 4.6*b*—*Edge enhancement by differentiation. Each element represents the direction of the gradient.*

The Laplacian operations, (4.10) and (4.11), can be computed by the convolution operation, (4.13), with the 3 by 3 filter matrices [12]:

$$[h_1]=\begin{pmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{pmatrix} \tag{4.15}$$

$$[h_2]=\begin{pmatrix} 0 & -1 & 0 \\ -1 & 5 & -1 \\ 0 & -1 & 0 \end{pmatrix} \tag{4.16}$$

Whenever the filter weight matrix exceeds a size of about 13 by 13 elements, filtering in the frequency domain, including the necessary Fourier transforms, is faster than direct convolution. However, frequency-domain filtering requires that the dimensions of the input image be a power of 2.

The subtle brightness variations that define the edges and texture of objects are important for subjective interpretation [13]. An enhancement of local contrast may be achieved by suppressing slow brightness variations, which tend to obscure the interesting details in an image. Slow brightness variations are composed of low spatial frequencies, whereas fine detail is represented by higher spatial frequencies. A filter that suppresses low spatial frequency components enhances local contrast. One frequently used filter [14] has the transfer function

$$H(u, v)=1-[1-H(0, 0)]KL\frac{\sin \pi Ku}{\pi Ku}\ \frac{\sin \pi Lv}{\pi Lv} \tag{4.17}$$

which is shown in figure 4.7 for one dimension. The variables K and L are the dimensions of the filter. For K and L on the order of 51 to 201, only the lowest spatial frequencies are removed. Smaller filter sizes (K, $L=3$ to 21) can be used for edge enhancement. Enhancement of features perpendicular to the row or column directions is possible with one-dimensional filters (K or $L=1$). However, distortions in the form of an enhancement in certain additional directions are introduced. In the spatial domain the enhanced image is efficiently computed by subtracting the average of a K by L area from the recorded image for each point. Figure 4.8 shows the effects of the filter described by equation (4.17) for $K=L=101$, 31, and 11. The larger the filter size, the smaller is the increase in amplitude at high spatial frequencies.

Enhancement of fine detail in visible images may be obtained by homomorphic filtering (see sec. 2.2.5) based on the illumination–reflectance model, equation (2.1), for image formation [15]. The illumination component is usually composed of low spatial frequencies, but the re-

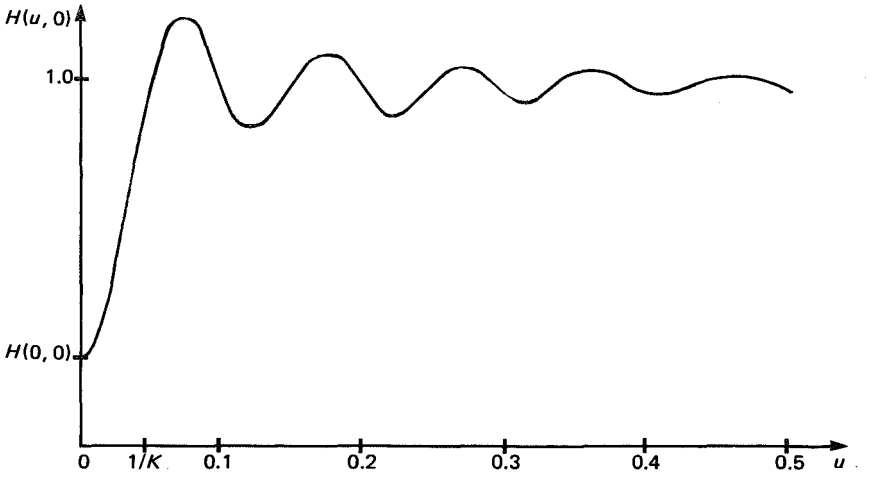


FIGURE 4.7—*Transfer function of local-contrast-enhancement filter (for K = 21).*

flectance component is characterized by high spatial frequencies, representing fine detail in the image. By applying the logarithmic transform

$$\ln f = \ln (ir) = \ln i + \ln r \tag{4.18}$$

the relation between illumination and reflectance becomes additive. Linear high-pass filtering may then be applied to suppress the illumination and to amplify the reflectance component relatively, thereby achieving enhancement of fine detail. A following exponentiation converts the filtered image back into the intensity domain. Figure 4.9 is a Landsat MSS 7 image enhanced by homomorphic filtering. An overall sharpening of an image can be achieved by application of a filter that compensates for the low-pass or blurring characteristics of the imaging system and that is adapted to the band-pass characteristics of the human visual system.

4.4 Color Enhancement

The human visual system can discriminate only about 20 to 30 shades of gray under a given adaptation level. Under the same conditions it discriminates a much larger number of color hues. Thus, the use of color provides a dramatic increase in the amount of information that can be displayed [16]. The color perceived by the observer is a result of the properties of the human visual system and of the display device used, i.e., the type of display medium in connection with the type of display excitation (e.g., film and primary-color illuminants) [17]. When display

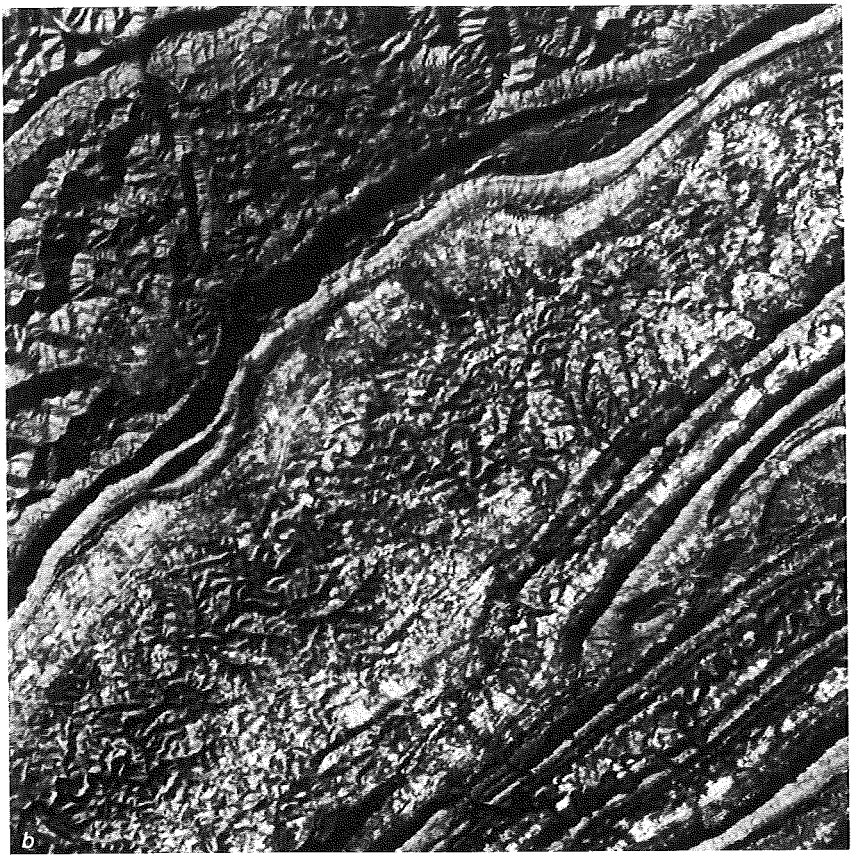


FIGURE 4.8a—*Recorded contrast-enhanced image (scene 692-15192).*

FIGURE 4.8*b*—*Image in part a enhanced with a 101 by 101 filter.*

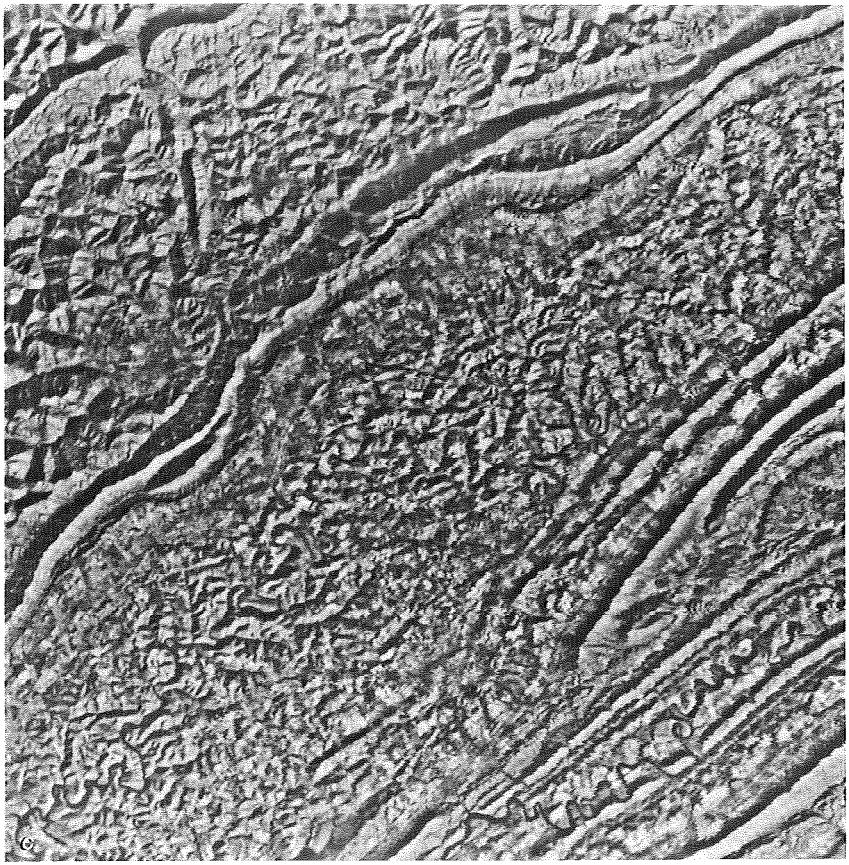


FIGURE 4.8c—*Image in part a enhanced with a 31 by 31 filter.*

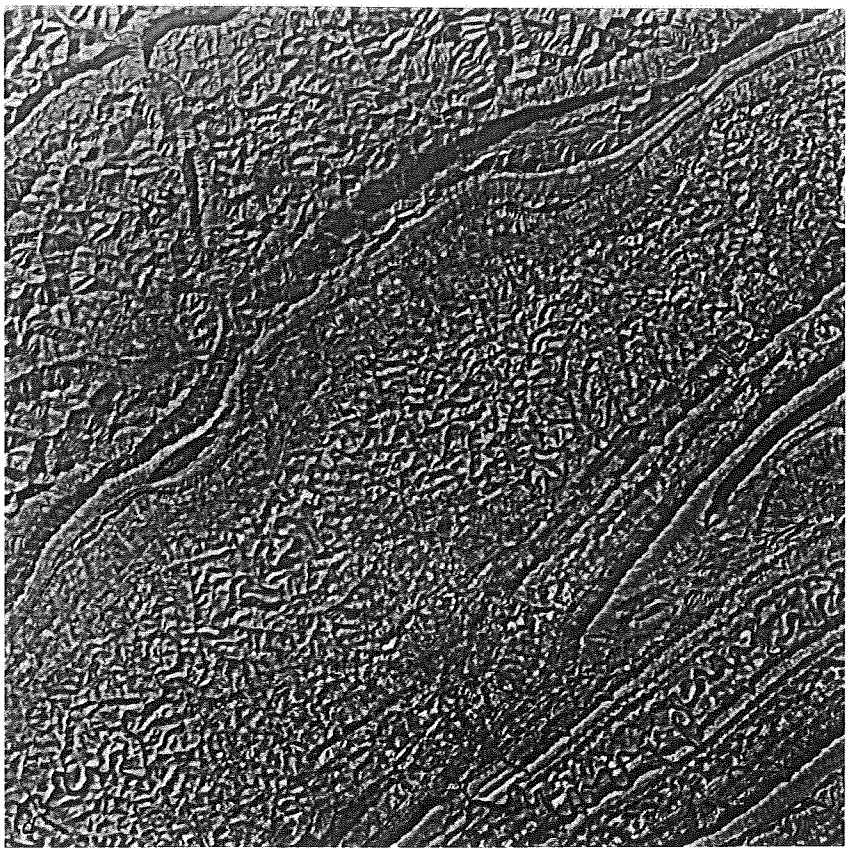


FIGURE 4.8*d*—*Image in part a enhanced with an 11 by 11 filter.*

FIGURE 4.9a—*Recorded contrast-enhanced image.*

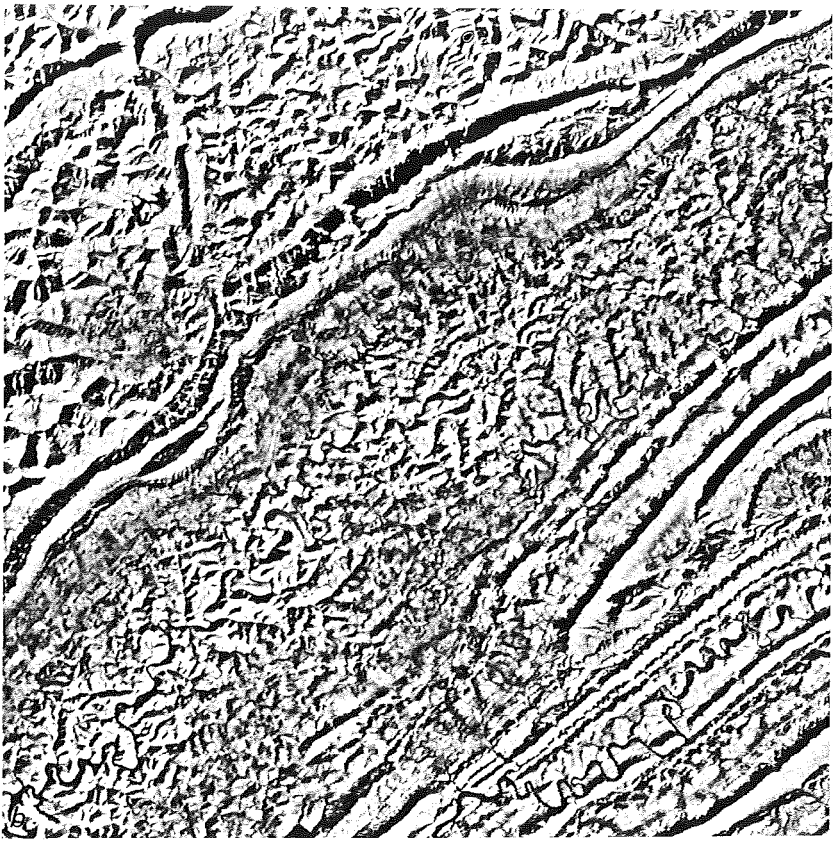


FIGURE 4.9*b*—*Result of high-emphasis filtering with a 31 by 31 filter after logarithmic transformation of image in part a.*

devices having slightly different sets of primary colors are employed, a transformation may be necessary to obtain color consistency (e.g., if an image is displayed temporarily on a television monitor and then recorded on film).

Digital multi-images may be displayed as color pictures by selecting three components for assignment to the primary colors. By varying the values of these components, all colors realizable within the constraints of the display medium may be generated. A color space, being linear in the color parameters brightness, hue, and saturation (see sec. 2.8.2), therefore in general does not lead to a visually perceived linear color range if linear relationships between color parameters and primary components are used. However, the components of a multi-image or the primary colors may be transformed to the color parameters brightness, hue, and saturation. An approximately equal color distribution may then be obtained by subsequent independent nonlinear transformations of each color parameter. This approach is justified because color perception cannot be as simply defined in an image as for isolated uniform areas (upon which color order systems are based) but must be defined for spatial changes (variegation) [4].

If f_n, $n=1, \ldots, P$ denotes the P given component images, hue H, saturation S, and brightness B images may be defined as

$$H(j,k) = \tan^{-1} \frac{\sum_{n=1}^{P} f_n(j,k) \sin \phi_n}{\sum_{n=1}^{P} f_n(j,k) \cos \phi_n} \tag{4.19}$$

$$S(j,k) = 1 - \frac{\min f_n(j,k)}{\max f_n(j,k)} \qquad n=1, \ldots, P \tag{4.20}$$

$$B(j,k) = \frac{\max f_n(j,k)}{K} \qquad \begin{matrix} n=1, \ldots, P \\ j=1, \ldots, M \\ k=1, \ldots, N \end{matrix} \tag{4.21}$$

and K is the maximum possible intensity value of the original component images. The angles ϕ_n determine the directions of the component image axis in a polar coordinate system (see fig. 2.27), and P is the dimensionality of the multi-range. For $P=3$, a choice of $\phi_1=0°$, $\phi_2=120°$, and $\phi_3=240°$ corresponds to three primary colors.

4.4.1 Pseudocolor

In observing black-and-white images, the eye responds only to brightness differences; i.e., black-and-white images restrict the operation of the

visual system to the vertical axis of the color perception space (fig. 2.27); the ability of the visual system to distinguish many hues and many saturations at each brightness level is not used [18]. By simultaneous brightness and chromatic variation, many more levels of detail can be distinguished. Small gray-scale differences in the black-and-white image that cannot be distinguished by the human eye are mapped into different colors. Consequently, more information can be extracted in a shorter time through the substitution of color for black and white.

The conversion of a black-and-white into a color image is achieved by a pseudocolor transformation. Three pseudocolor component images are produced by controlling the relationship between the colors in the final image and the corresponding intensity in the original. Proper choice of this relationship permits full use of the abilities of the human visual system to use hue and saturation in addition to brightness for the purpose of discrimination. The three component images are then combined by using appropriate amounts of three primary colors.

One pseudocolor technique is known as level slicing, where each gray level or a range of gray levels is mapped into a different color. To avoid the introduction of artificial contours, as in level slicing, a continuous transformation of the gray scale into the color space may be performed [4]. One transformation that results in a maximum number of discernible levels is to project the gray values onto a scale of hues. The projection can be scaled and shifted to include only a particular part of the entire hue scale.

Pseudocolor enhancement is illustrated in figure 4.10, showing a Heat Capacity Mapping Mission (HCMM) thermal infrared image (spectral band, 10.5 to 12.5 μm; spatial resolution, 500 m) of the Eastern United States. Figure 4.10*a* is a contrast-enhanced black-and-white image. Figure 4.10*b* shows the corresponding level-sliced pseudocolor image with the intensity range divided into 32 colors. Blue represents the coldest, and red and white represent the warmest, areas in the thermal infrared image. Figure 4.10*c* shows the pseudocolor image obtained by mapping the gray scale of the black-and-white image onto the hue scale. The transformation can be designed to obtain an approximately equal visual distribution of colors produced with a particular display device.

4.4.2 False Color

False-color enhancement is used to display multispectral information from the original scene where the spectral bands are not restricted to the visible spectrum. The goal is to present certain spectral information from the object scene rather than to have color fidelity. By assuming spatial registration, any three components of a multispectral image may be selected and combined by using appropriate primary colors. Variations in the spectral response of patterns then appear as color differences in the

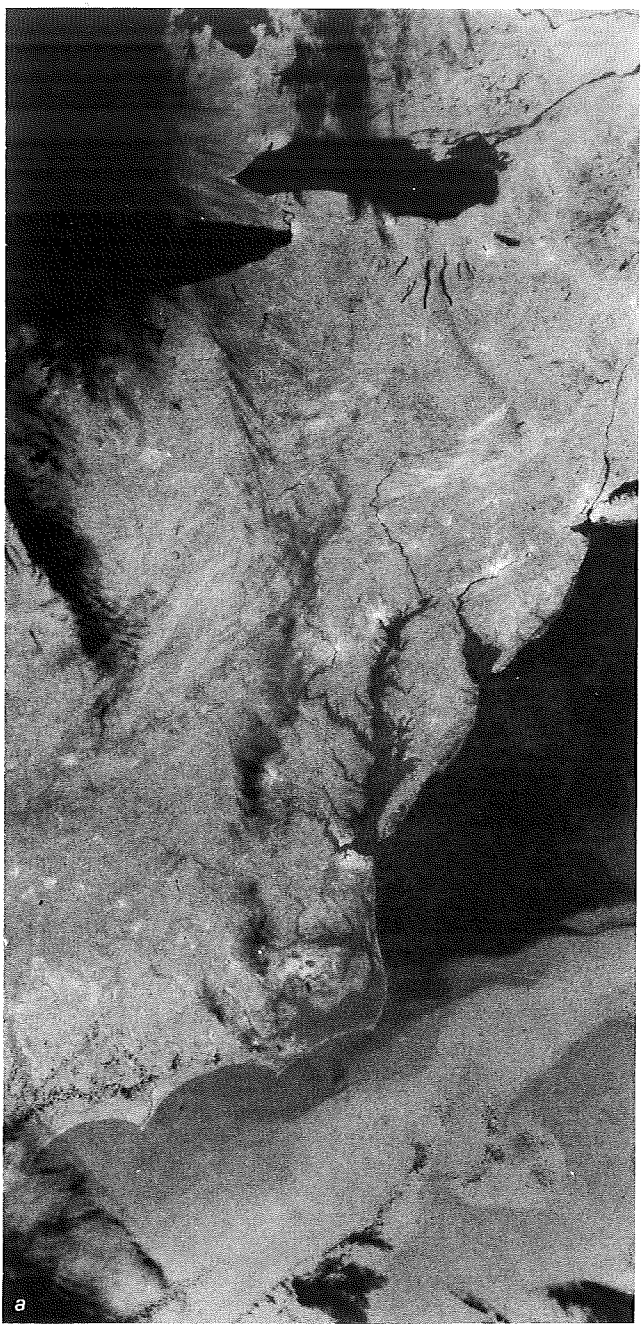


FIGURE 4.10*a*—*Contrast-enhanced black-and-white image.*

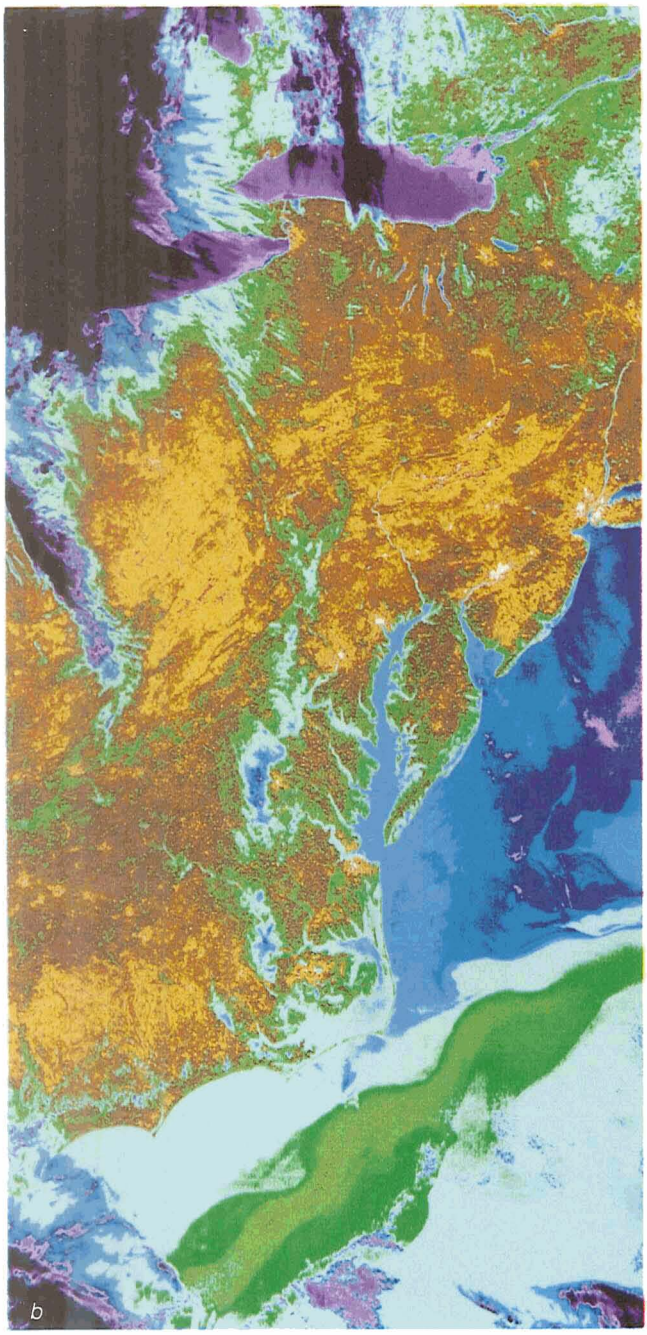


FIGURE 4.10*b*—*Level-sliced pseudocolor version of image in part a with 32 colors.*

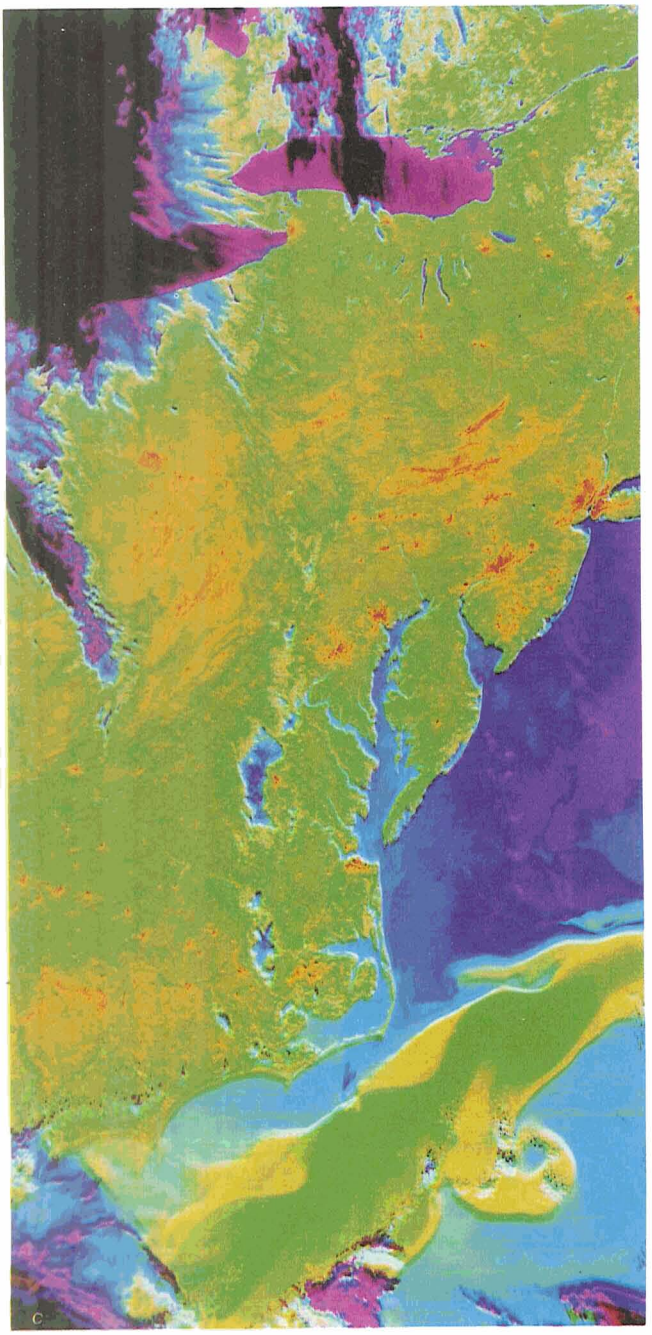


FIGURE 4.10c—*Gray scale of image in part a mapped onto hue scale in color space.*

composite image. These colors may show no similarity with the actual colors of the pattern. Ratios, differences, and other transformations of the spectral bands may also be displayed as false-color composites.

Producing a good false-color image requires careful contrast enhancement of each component to obtain a good balance and range of colors in the composite. Generally, good results are obtained by applying contrast transformation to the three component images in such a way that their histograms look similar in shape and that each individual component has appropriate contrast when displayed as a black-and-white image. These transformations insure good color and brightness variations. They can be performed by automatic histogram normalization or by individual determination of contrast characteristics from the original histograms. Histogram flattening by approximation of a ramp cumulative distribution function of the gray values often tends to produce high saturation and excessive contrast. Depending on the scene, this normalization may be desirable or not. The approximation of a normally distributed (Gaussian) histogram produces less saturation. The transformations should assign the mean of each enhanced component to the center of the dynamic range of the display device.

Filtering of the component images may be required before false-color composition. Some filtering techniques, such as edge enhancement to correct for the low-pass characteristics of the imaging system or band-pass filtering to enhance visual perception, can be performed separately on the component images without loss of color information. The false-color image pair in figure 4.11 shows the result of filtering three Landsat MSS image components with a logarithmic band-pass filter adapted to the human visual system. The elimination of large-scale brightness variations by edge enhancement with the local-contrast-enhancement filter and the transfer function in equation (4.17), however, results in a loss of color information. This loss occurs because the average brightness of any homogeneous region in the image whose size is of the order of the filter size is zero. Therefore, a color composite of images enhanced with smaller filter sizes has a grayish appearance. This effect is illustrated in figure 4.12*a*, which shows a false-color display of three Landsat MSS components, each filtered with the edge-enhancement filter of equation (4.17), with $K=L=31$. The color loss is less obvious in the homomorphically filtered version of the same scene shown in figure 4.12*b*.

This problem can be avoided by separating the color information from brightness and filtering only the brightness component. The original component images are transformed to the hue, saturation, and brightness color coordinate system, and filtering is performed only on the brightness component. The inverse transformation is then applied before display. Furthermore, more than three component images may be transformed to the color space. Figure 4.13 shows the result of transforming four

FIGURE 4.11—*Color edge enhancement by band-pass filtering. (a) False-color display of three contrast-enhanced Landsat images (MSS 4 = blue; MSS 5 = green; MSS 7 = red). (b) False-color display of the edge-enhanced Landsat MSS image components.*

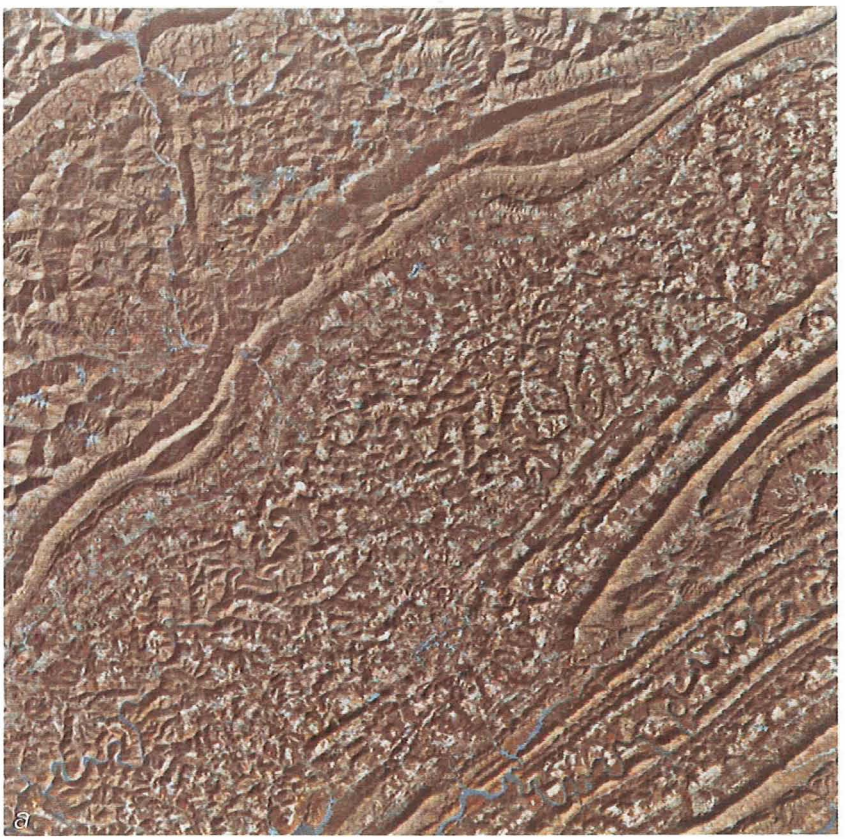


FIGURE 4.12a—*Color edge enhancement with a 31 by 31 filter applied directly to the image components.*

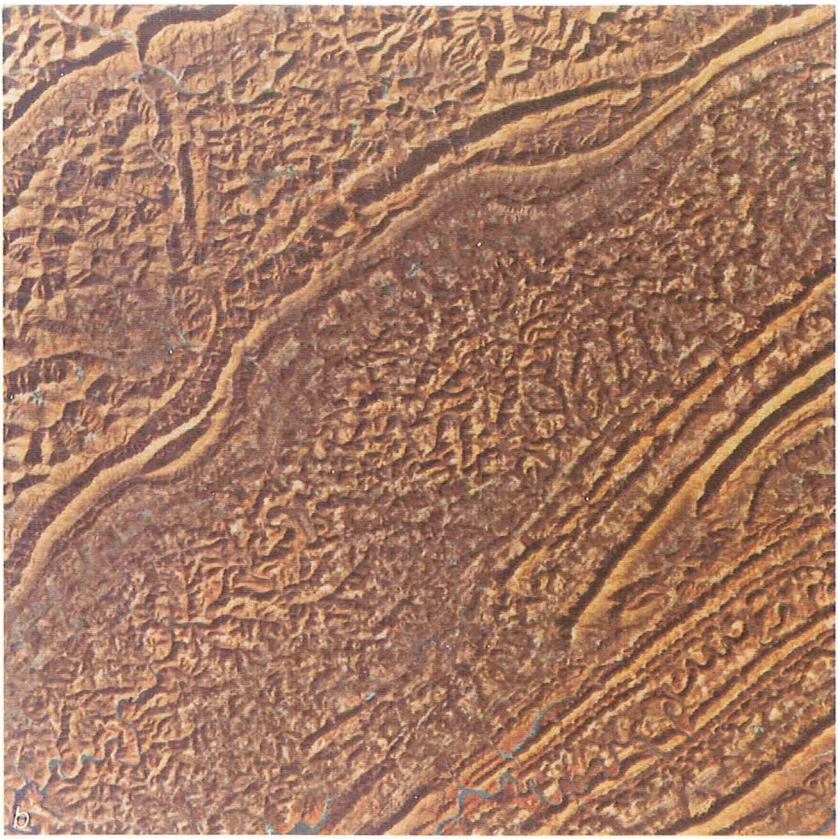


FIGURE 4.12*b*—*Color edge enhancement by homomorphic filtering of the image components with a 31 by 31 filter.*

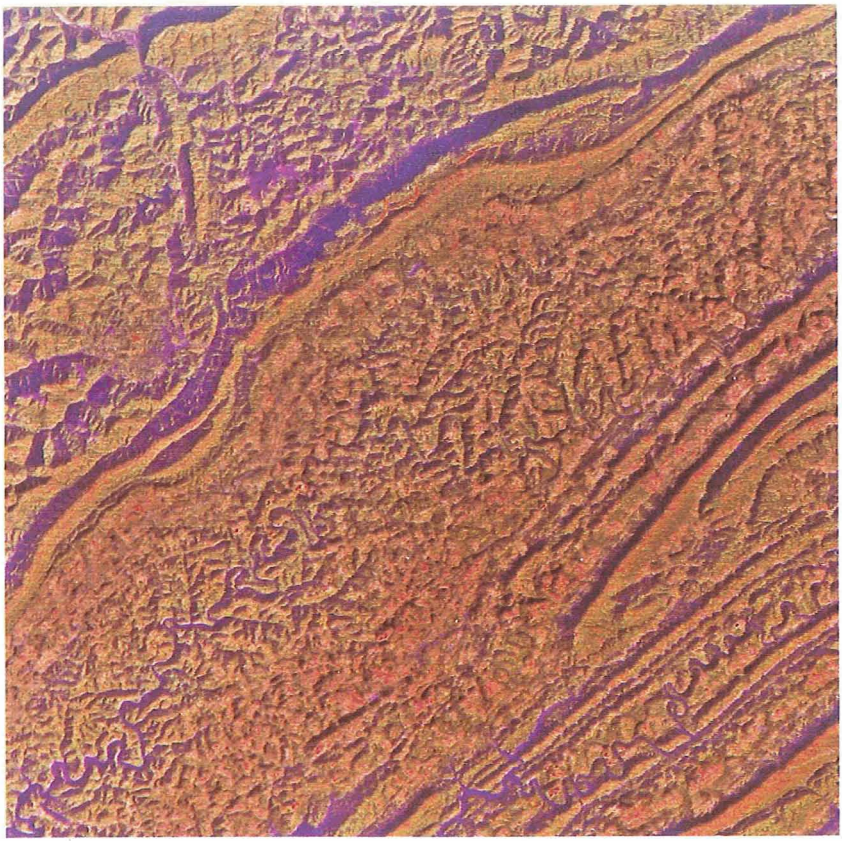


FIGURE 4.13—*Color edge enhancement by filtering in color space. Brightness component was filtered with a 31 by 31 edge-enhancement filter with characteristics given by equation (4.17).*

Landsat MSS bands with equations (4.19), (4.20), and (4.21) and performing edge-enhancement filtering on the brightness component.

4.5 Multi-Image Enhancement

Multi-images convey more information than monochrome images. Multi-images are obtained by imaging a scene in more than one spectral band or by monitoring a scene over a period of time. Multi-image enhancement techniques involve independent contrast enhancement of the component images or linear and nonlinear combinations of the component images, including ratioing, differencing, and principal-component analysis. The enhanced components may be displayed as false-color composites.

4.5.1 Ratioing

Multispectral images may be enhanced by ratioing individual spectral components and then displaying the various ratios as color composites. Ratioing two spectral component images suppresses brightness variations due to topographic relief and enhances subtle spectral (color) variations [19].

If $\mathbf{g}$ is a multi-image with N components g_i, $i=1, \ldots, P$, then a ratio image $g_k{}^R$, $k=1, \ldots, P(P-1)$, is given by

$$g_k{}^R = a\frac{g_i}{g_j} + b \tag{4.22}$$

The contrast in ratio pictures is greater for features with ratios larger than unity than for those with ratios less than unity. By computing the logarithm of the ratios, equal changes in the denominator and numerator pictures result in equal changes in the logarithmic ratio image. Thus, the logarithmic ratio image shows greater average contrast between features.

Ratioing also enhances random noise or coherent noise that is not correlated in the component images. Thus, striping should be removed before ratioing. (See sec. 3.2.3.) Atmospheric effects may also be enhanced by ratioing. The diffuse scattered light from the sky assumes a larger portion of the total illumination as the incident angle of direct color illumination decreases. The effect is that the color of the scene is partly a function of topography. The scattered light from the sky can be estimated by examination of dark features shaded from the Sun by large clouds. The resulting values s_i represent the scanner readings that would occur if the scene were illuminated only by light scattered from the sky. (See sec. 3.2.2.) Because these values do not change significantly over a scene of limited size, a first-order atmospheric correction for ratioing may be performed by

$$g_k{}^R = a\frac{g_i - s_i}{g_j - s_j} + b \tag{4.23}$$

The selection of the most useful ratios and their combination into color composites is a problem. The number of possible ratios from a multi-image with P components is $n=P(P-1)$. The number of possible combinations of three of these ratios into a color composite is $m=n!/[3!\ (n-3)!]$. The primary colors may be assigned to each triplet in six different ways. Thus, ratioing is only efficient when a priori knowledge of the useful ratios and color combinations is available.

Ratioing has been successfully applied for geologic applications [20]. False-color composites of ratio images provide the geologist with information that can not be obtained from unprocessed images. Figure 4.14*a* shows a false-color composite of Landsat MSS bands 4, 5, and 7 of a geologically interesting area in Saudi Arabia. Figure 4.14*b* is a false-color composite of the ratio images of MSS bands 4 to 5, 5 to 6, and 6 to 7, where each ratio image was contrast enhanced by histogram flattening. Figure 4.15 compares the results of contrast enhancement and ratioing. Figure 4.15*a* represents a linear contrast enhancement of a Landsat MSS false-color image of the Sahl al Matran area in Saudi Arabia. Figure 4.15*b* is a false-color composite of the nonlinearly enhanced image components obtained by histogram flattening. Figure 4.15*c* shows the false-color composite of the contrast-enhanced ratio images.

4.5.2 Differencing

Temporal changes or spectral differences may be enhanced by subtracting components of a multi-image from each other. As for ratioing, the component images must be spatially registered. Temporal changes are determined by subtracting two component images taken at different times. Spectral differences are determined by subtracting two component images taken in different wavelength bands. This difference is a representation of the slope of the spectral reflectance curve if the images are corrected for distorting factors, such as instrument calibration and atmospheric transmittance.

Differences may have positive or negative values. For display, the differences must be scaled to lie within the dynamic range of the display device. The scaling should also produce sufficient contrast to make changes visible. These requirements are fulfilled by

$$g_k^D = a(g_i - g_j) + b \tag{4.24}$$

where g_i and g_j are component images and g_k^D is the difference image. The constants a and b are usually determined so that zero difference is represented as midscale gray ($g_k^D=128$ for 8-bit quantization), and differences of magnitude greater than 64 are saturated to white for positive differences and black for negative differences. Small differences are best displayed by pseudocolor enhancement.

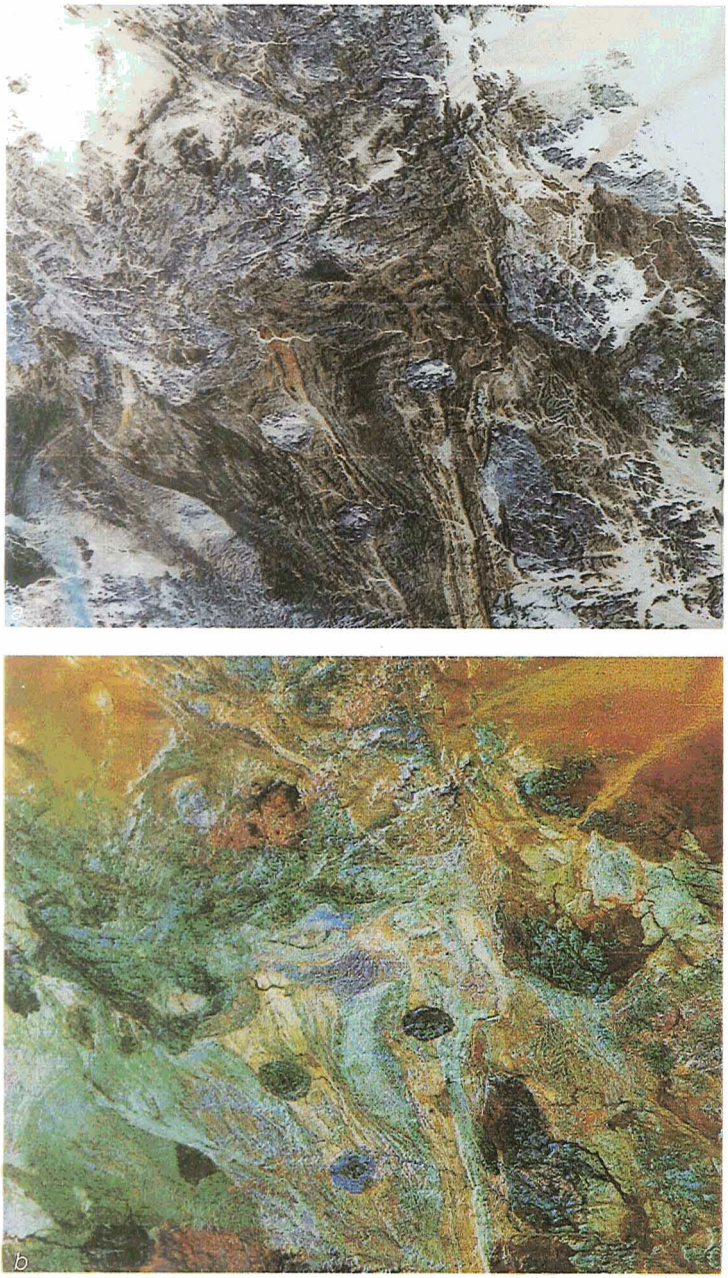


FIGURE 4.14—*Color display of ratio images. (a) False-color composite of Landsat MSS bands 4 (blue), 5 (green), and 7 (red). (b) False-color composite of ratio images of MSS bands 4 to 5 (blue), 5 to 6 (green), and 6 to 7 (red) (scene 1226-07011).*

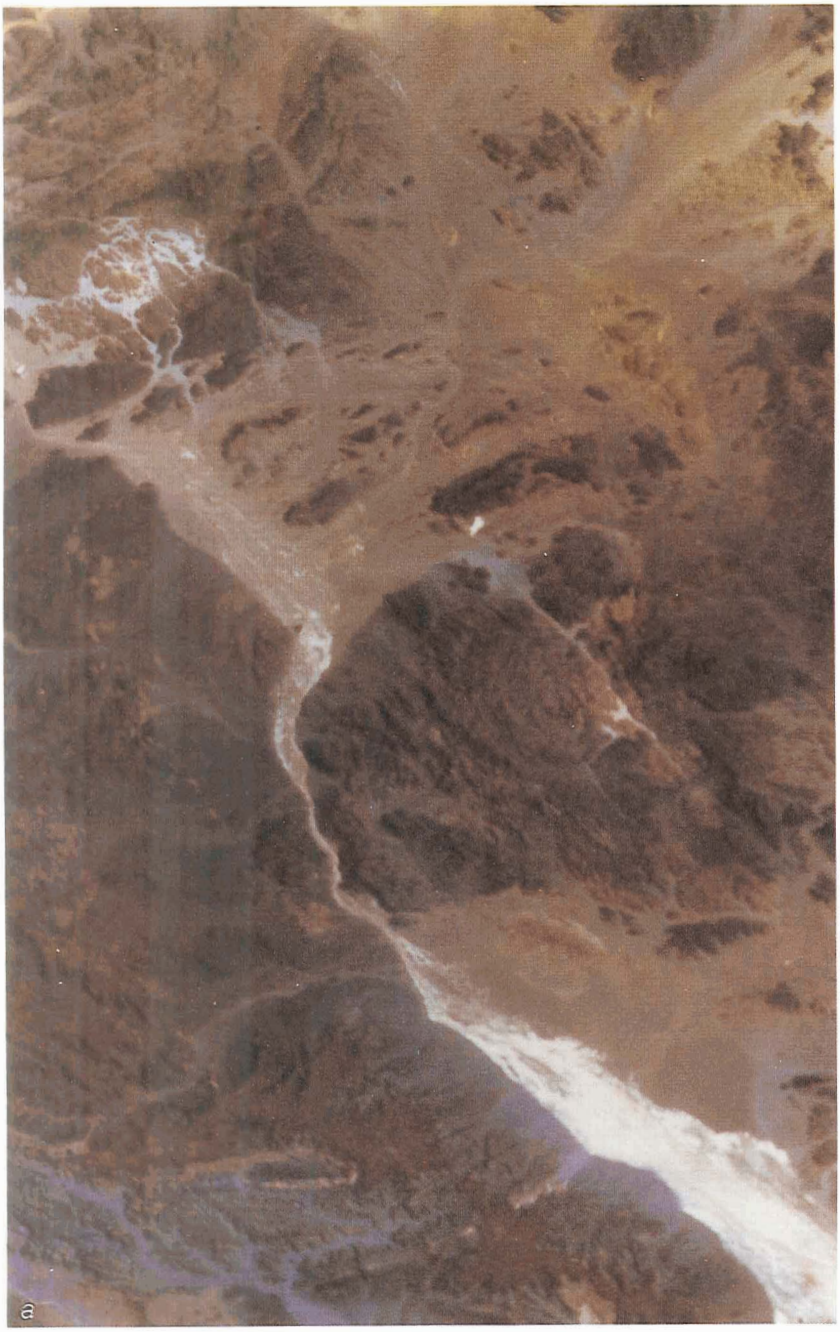


FIGURE 4.15—*Contrast enhancement and ratioing (scene 1357-07285).*

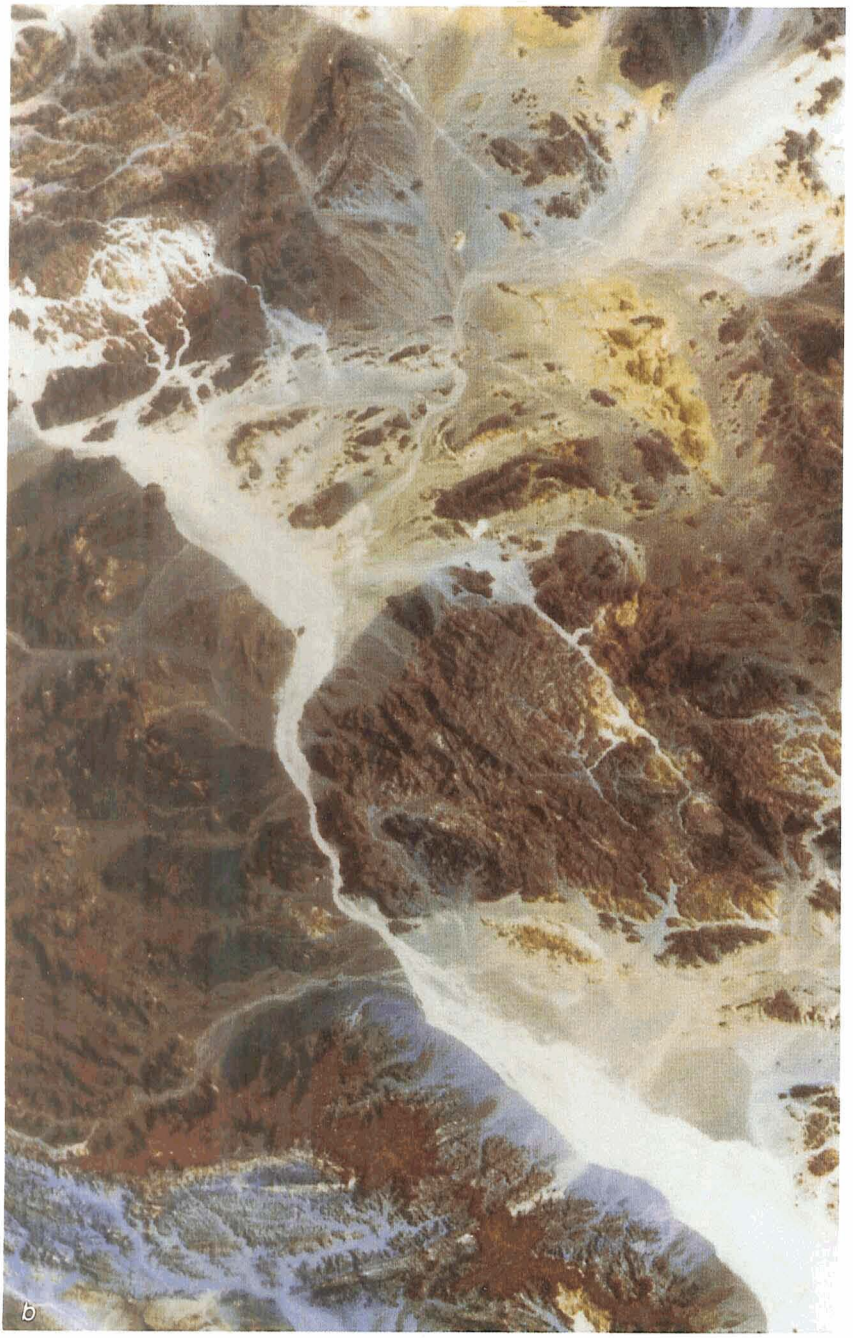


FIGURE 4.15—*Continued.*

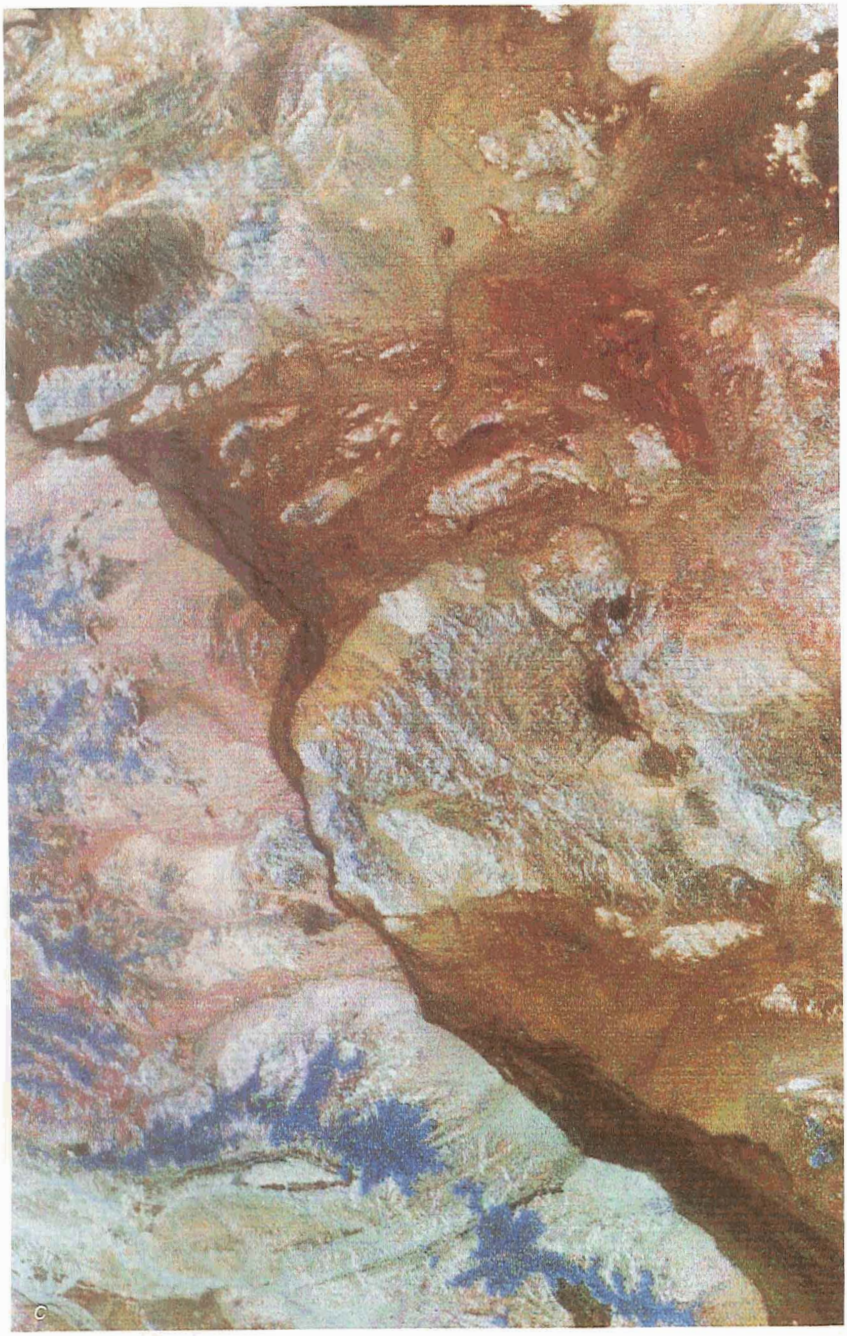


FIGURE 4.15—*Continued.*

Differencing is a simple method for edge enhancement. Shifting an image by one row and by one column and subtracting it from the original produces a picture that represents first differences in the row and column directions.

An application of image differencing for the detection and enhancement of temporal change is illustrated in figure 4.16 [21]. Figures 4.16*a* and 4.16*b* show two registered Landsat MSS false-color images taken at different times. Figure 4.16*c* is a display of the differences between these two images. Areas of major change appear light (increased reflectance). Areas with little change are displayed in midgray. To extract the areas of major change, the difference image is brought to a threshold (see sec. 7.2) for values three standard deviations above and below the mean. The resulting binary image representing the changed areas in white is shown in figure 4.16*d*. Differencing can only detect the location and size of changed areas. The type of change may be determined by image classification. (See ch. 8.) The binary image in figure 4.16*d* may be used as a mask to extract the changed areas for classification, thereby considerably reducing the amount of data to be processed [21]. The total change combined with figure 4.16*b* is shown in figure 4.17*a*. The classified changed areas are overlaid on one black-and-white component (MSS 5) of the multi-image in figure 4.17*b*. Changes in agricultural and in urban and industrial areas are shown in green and red, respectively. The yellow areas represent changes that could not be uniquely identified.

4.5.3 Transformation to Principal Components

The problem of selecting a subset of a multi-image for enhancement by false-color compositing, ratioing, or differencing is generally difficult. Usually an intuitive selection, based on known physical characteristics of the scenes and on experience, is made. Multispectral images often exhibit high correlations between spectral bands; therefore the redundancy between the components of such multi-images may be significant.

The Karhunen–Loéve (K–L) transform (see sec. 2.6.1.4) to principal components provides a new set of component images that are uncorrelated and are ranked so that each component has variance less than the previous component. Thus, the K–L transform can be used to reduce the number of spectral components to fewer principal components that account for all but a negligible part of the variance in the original multispectral image. The principal component images may be enhanced and combined into false-color composites.

The principal component image $\mathbf{g}^c$ is obtained from the original P components of a multi-image $\mathbf{g}$ by the transformation

$$\mathbf{g}^c = \mathbf{T}(\mathbf{g} - \mathbf{m}) \tag{4.25}$$

FIGURE 4.16—*Detection of temporal change by differencing.*

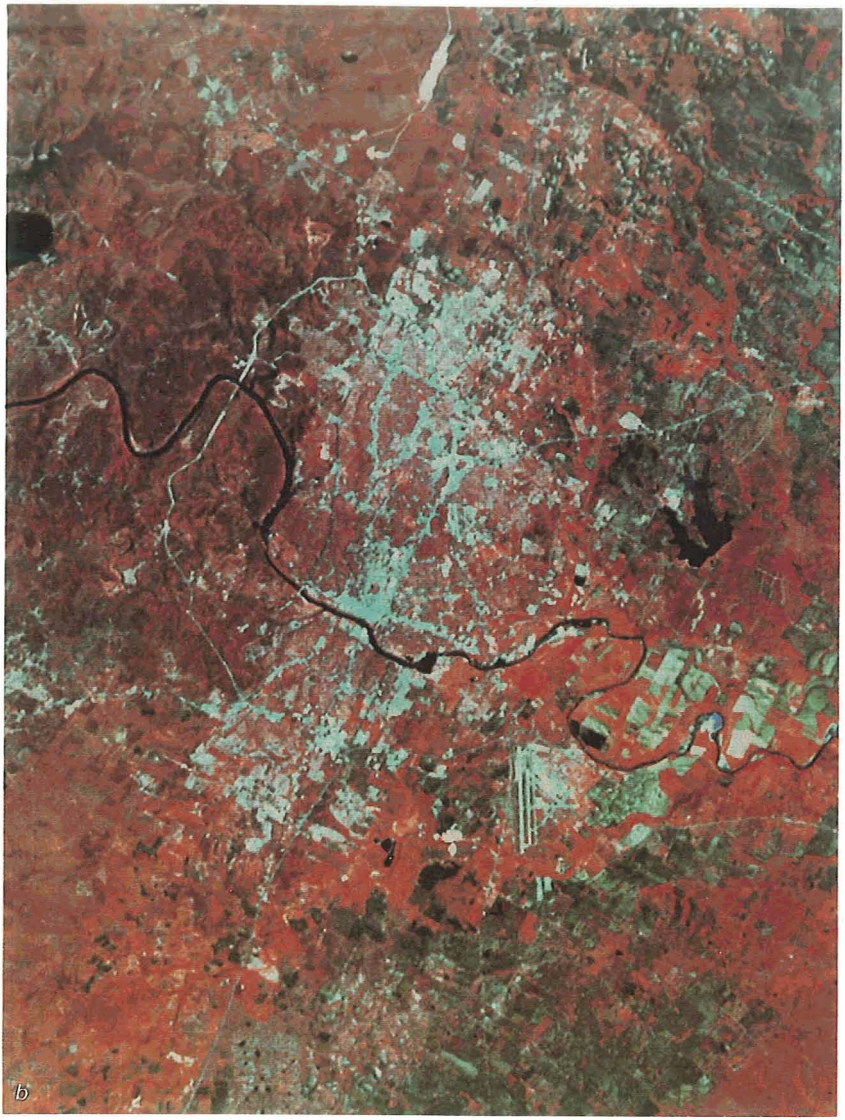


FIGURE 4.16—*Continued.*

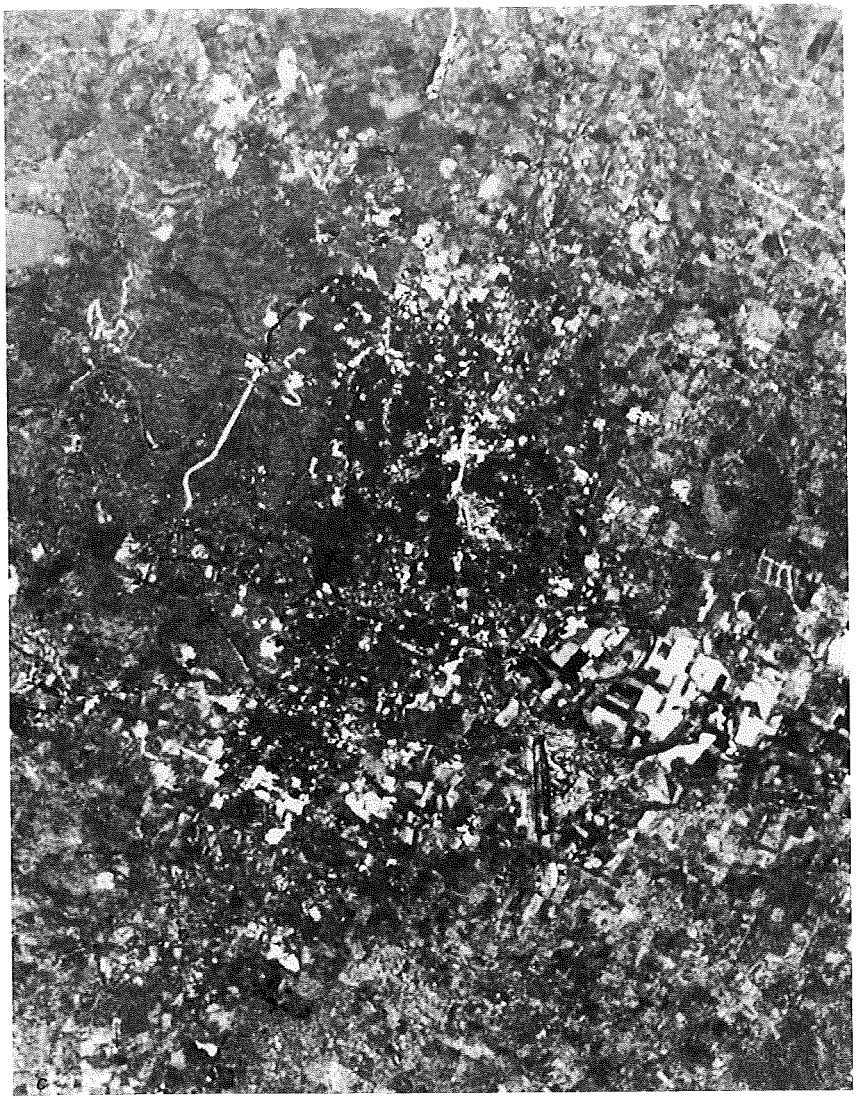


FIGURE 4.16—*Continued.*

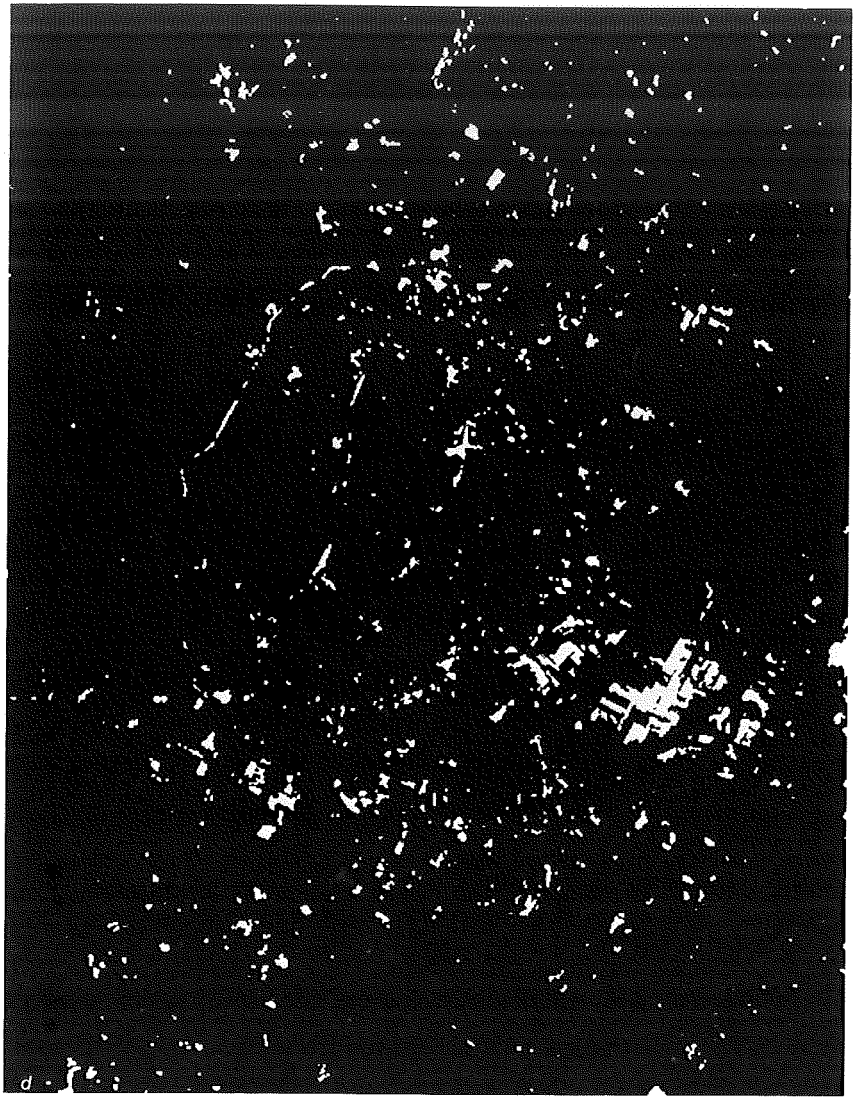


FIGURE 4.16—*Continued.*

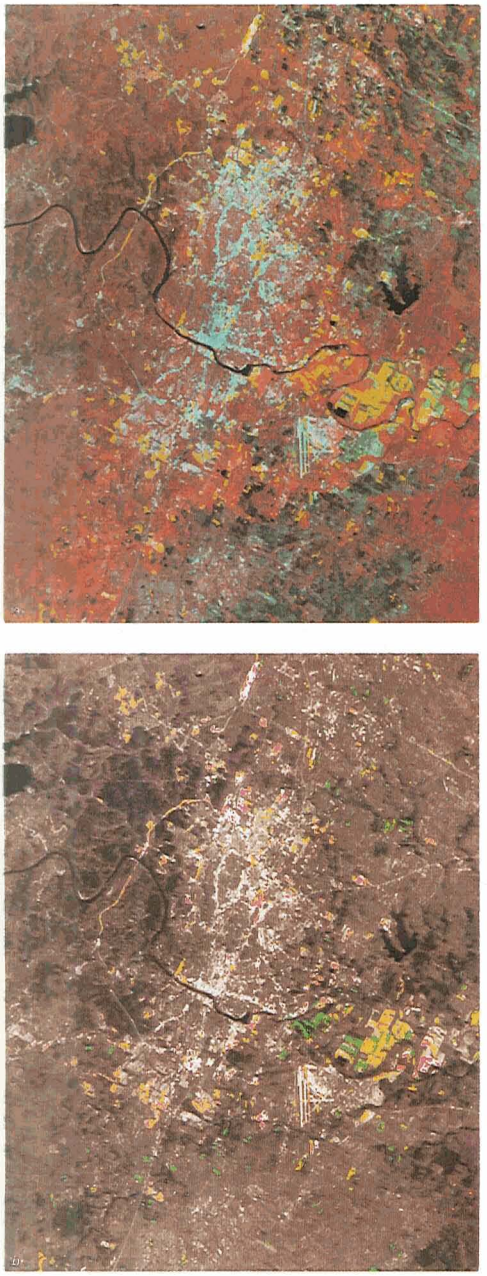


FIGURE 4.17—*Combination of changed areas with original image. (a) Total change overlaid on image in figure 4.16b. Total change is shown in yellow. (b) Classified change overlaid on MSS band 5 (agriculture = green; urban and industrial areas = red; ambiguous areas = yellow). (Images courtesy of R. McKinney, Computer Sciences Corp.)*

where $\mathbf{g}$ is a vector whose elements are the components at a given location (j, k) in the original multi-image, $\mathbf{m}$ is the mean vector of $\mathbf{g}$; i.e., $\mathbf{m}=\mathbf{E}(\mathbf{g})$. The components of vector $\mathbf{g}^c$ are the principal components at the location (j, k), $\mathbf{T}$ is the P by P unitary matrix whose rows are the normalized eigenvectors $\mathbf{t}_p$, $p=1, \ldots, P$ of the spectral covariance matrix $\mathbf{C}$ of $\mathbf{g}$ arranged in a descending order according to the magnitude of their corresponding eigenvalues: $\mathbf{T}=(\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_P{}^T.)$

The covariance matrix is computed as:

$$\mathbf{C}=E\{(\mathbf{g}-\mathbf{m})(\mathbf{g}-\mathbf{m})^T\} \tag{4.26}$$

The eigenvalues λ_p and the eigenvectors $\mathbf{t}_p$ of $\mathbf{C}$ are determined by solving the equation (see sec. 2.6.1.4)

$$\mathbf{C}\mathbf{t}_p=\lambda_p\mathbf{t}_p \tag{2.123}$$

The eigenvectors $\mathbf{t}_p$ form the basis of a space in which the covariance matrix is diagonal. Therefore, the principal components are uncorrelated. The eigenvalues λ_p are the variances $\sigma_p{}^2$ of the principal components; i.e., $\sigma_p{}^2=\lambda_p$, $p=1, \ldots, P$. They indicate the relative importance of each component.

The procedure to use the K–L transform for the enhancement of multi-images consists of the following steps:

1. Compute the covariance matrix equation (4.26) of the multi-image and its eigenvectors in the spectral dimension (eq. 2.123).
2. Transform the given image vector $\mathbf{g}$ to principal components by using equation (4.25).
3. For false-color display, select and enhance the contrast of three components and combine them into a color composite.

The data that enter into the computation of the covariance matrix determine the characteristics of the enhancement. Computing the covariance matrix from all data of a scene results in a mean enhancement. To discriminate patterns of interest in the final display, the covariance matrix must be computed for selected training areas representing these patterns. An enhancement of the signal with respect to additive uncorrelated noise is achieved by the K–L transform. Let $\mathbf{g}$ be given by

$$\mathbf{g}=\mathbf{f}+\mathbf{n} \tag{4.27}$$

where the elements of the noise vector $\mathbf{n}$ are assumed to be uncorrelated, zero-mean, identically distributed random variables. Thus, the noise covariance matrix is $\mathbf{C}_n=\sigma_n{}^2\mathbf{I}$, where $\mathbf{I}$ is the P by P identity matrix. The covariance matrix of $\mathbf{g}$ is

$$\mathbf{C}=\mathbf{C}_t+\mathbf{C}_n \tag{4.28}$$

and equation (2.129) becomes

$$(\mathbf{C}_f+\mathbf{C}_n)\mathbf{t}_p=\lambda_p\mathbf{t}_p \tag{4.29}$$

or

$$\mathbf{C}_f\mathbf{t}_p=(\lambda_p-\sigma_n{}^2)\mathbf{t}_p=\lambda_p{}^f\mathbf{t}_p \tag{4.30}$$

Thus, the eigenvectors forming the transformation matrix $\mathbf{T}$ are insensitive to the noise, and the eigenvalues of the noise-free image $\mathbf{f}$ are $\lambda_P{}^{\mathbf{f}}$, where

$$\lambda_P{}^{\mathbf{f}}=\lambda_p-\sigma_n{}^2 \tag{4.31}$$

The transformation (4.25) has no effect on the variance of uncorrelated, identically distributed noise. The maximum signal-to-noise ratio (SNR) in the original multi-image may be defined as [22]:

$$\mathrm{SNR}_{\mathbf{f}}=\frac{\sigma_{\mathbf{f}}{}^2}{\sigma_n{}^2} \tag{4.32}$$

where

$$\sigma_{\mathbf{f}}{}^2=\{\sigma_{f_1}{}^2,\ \sigma_{f_2}{}^2,\ \ldots,\ \sigma_{f_P}{}^2\} \tag{4.33}$$

The maximum SNR in the principal component image is

$$\mathrm{SNR}_{g^c}=\frac{\lambda_1}{\sigma_n{}^2} \tag{4.34}$$

Because $\lambda_1>\sigma_{\mathbf{f}}{}^2$, an enhancement of the signal with respect to the noise is achieved.

The K–L transform permits estimation of the noise level in correlated multi-images. Because $\lambda_P{}^{\mathbf{f}}\sim 0$ for correlated data, equation (4.31) yields

$$\sigma_n{}^2\sim\lambda_P \tag{4.35}$$

Thus, the variances of the original components can be divided by the eigenvalue of the last principal component to give a measure of the SNRs in the original image.

Figure 4.18 shows eight components of a multispectral image taken by an airborne Multispectral Scanner Data System (MSDS) [23] over an area in Utah. The scanner resolution is approximately 6 m, and the data are quantized to 8 bits. The spectral channels, the corresponding wavelength bands, and the means and variances are shown in table 4.1. Each pixel in the multi-spectral image is a P-dimensional vector $\mathbf{g}$, with $P=8$, and the components are the quantized spectral intensities for that point. The covariance matrix associated with $\mathbf{g}$ is computed by equation (4.26). Its eigenvalues, the percentage variances in the principal components, and the cumulative percentage variances are shown in table 4.2. The principal component images $\mathbf{g}^c$ obtained by equation (4.25) are shown in figure 4.19. The first three principal component images contain 97 percent of the original data variance. The number of possible combinations of three of the original spectral components for false-color display is $n=P!/3!(P-3)!=56$. False-color composites of principal component images are shown in figure 4.20.

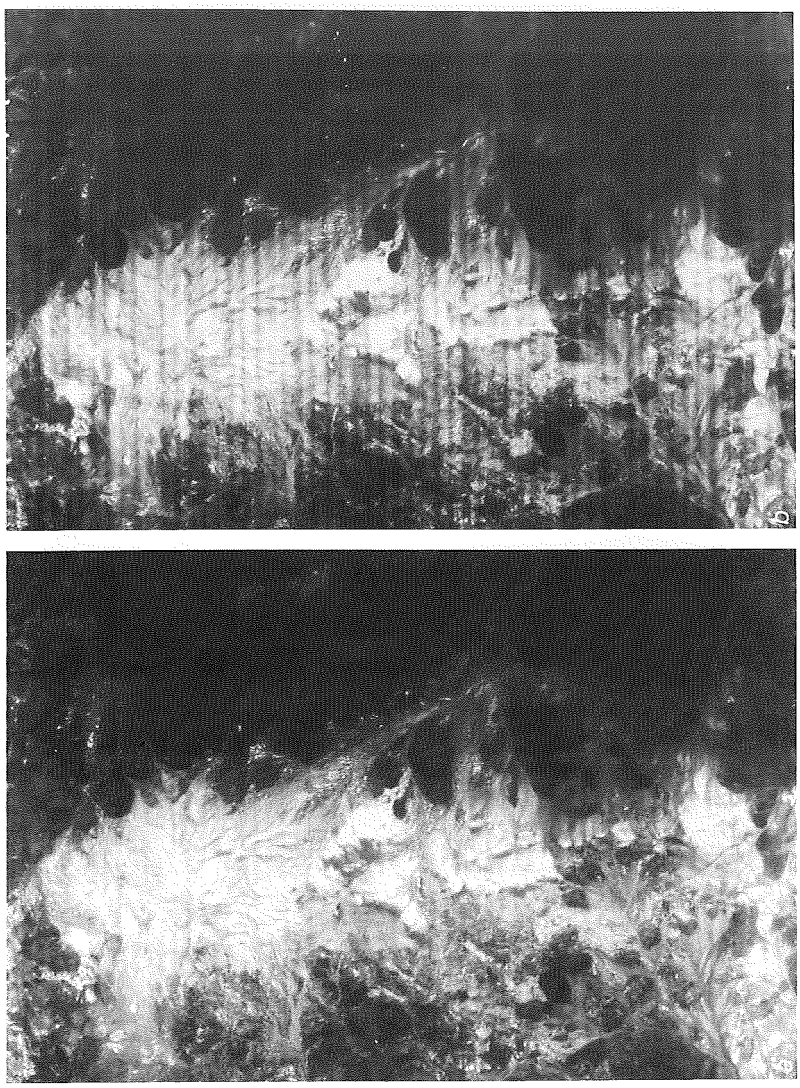


FIGURE 4.18—*Channel 1 (a) and 2 (b) components of a multispectral aircraft scanner image.*

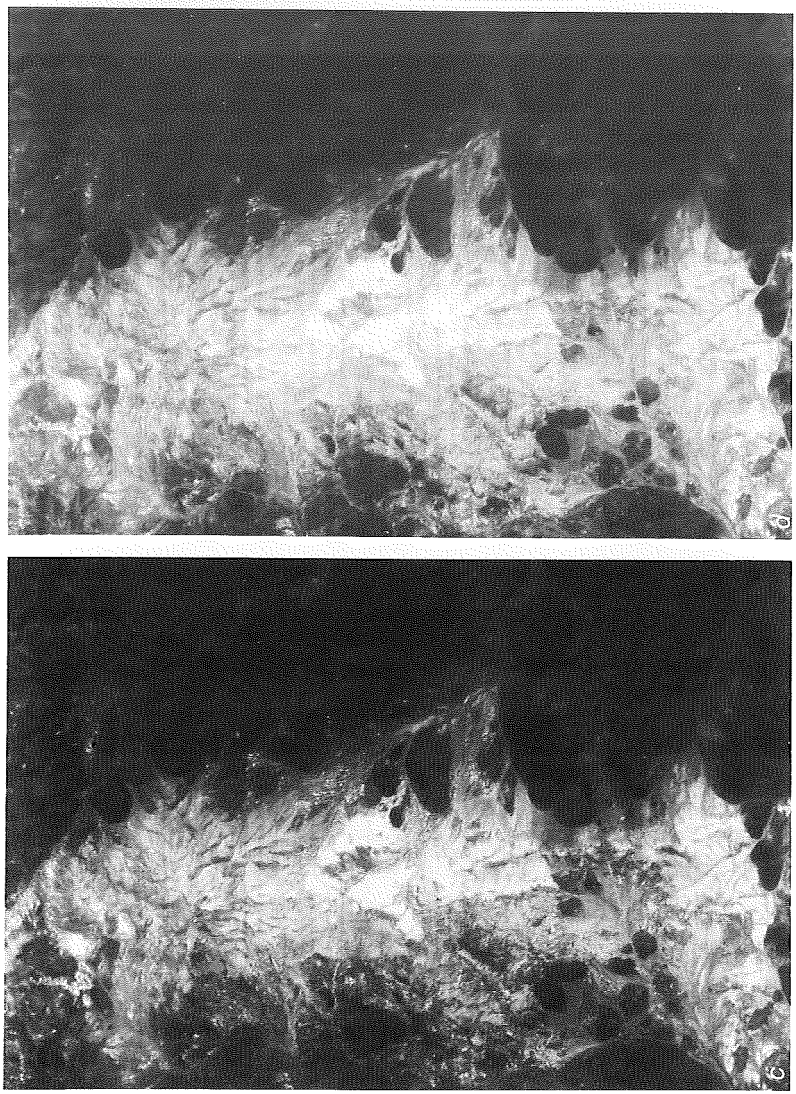


FIGURE 4.18—*Continued. Channel 3 (c) and 4 (d) components of a multispectral aircraft scanner image.*

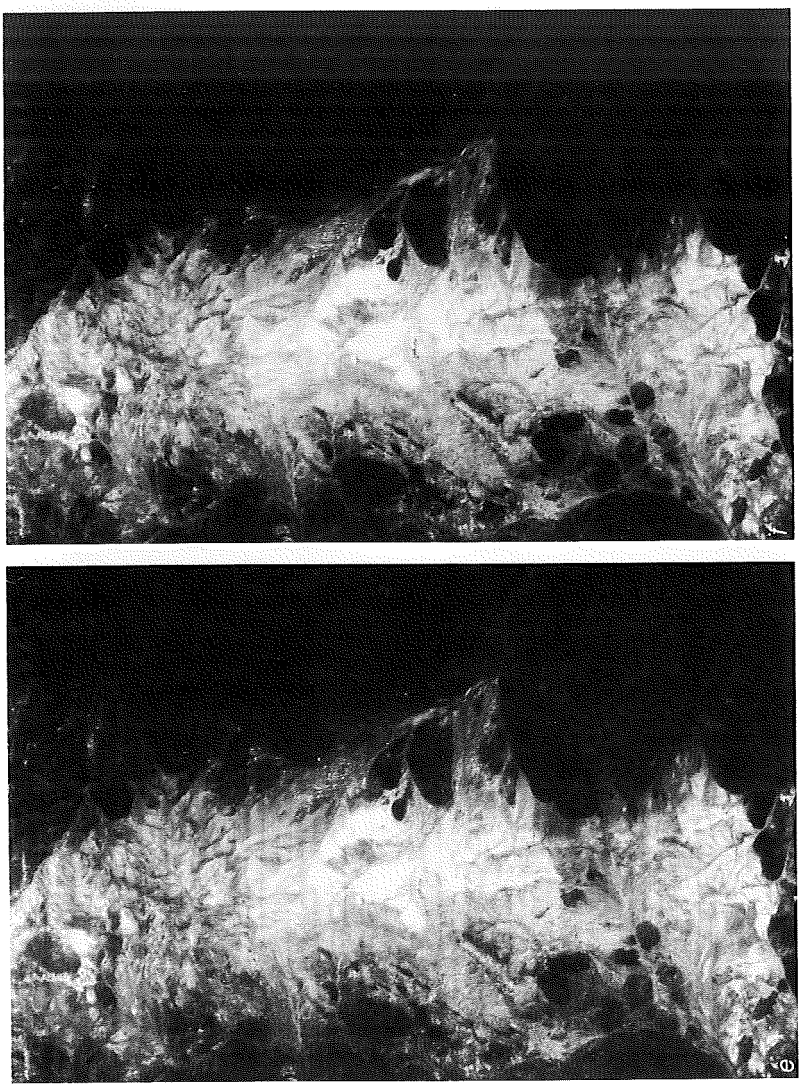


FIGURE 4.18—*Continued. Channel 5 (e) and 6 (f) components of a multispectral aircraft scanner image.*

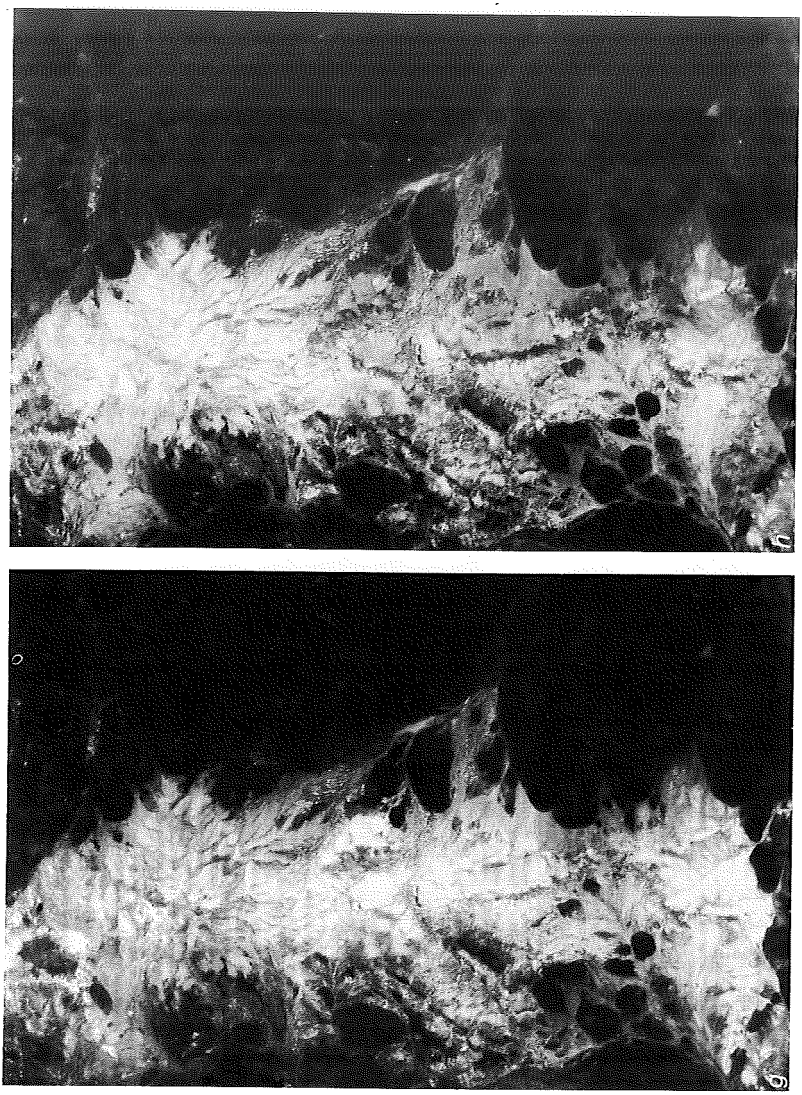


FIGURE 4.18—*Continued. Channel 7 (g) and 8 (h) components of a multispectral aircraft scanner image.*

TABLE 4.1—*Wavelength Bands, Mean Values, and Variances of Eight-Channel MSDS Multispectral Image*

Channel	Wavelength band (μm)	Mean	Variance
1	0.46–0.49	116.9	440.6
2	0.53–0.58	99.0	510.3
3	0.65–0.69	129.9	739.3
4	0.72–0.76	152.6	1,397.3
5	0.77–0.81	160.2	1,685.9
6	0.82–0.88	123.2	1,255.0
7	1.53–1.62	128.4	679.1
8	2.30–2.43	140.2	431.8

TABLE 4.2—*Eigenvalues and Percent Variances of Principal Components*

	Principal component							
Parameter	1	2	3	4	5	6	7	8
Eigenvalue	6090.1	472.8	342.7	119.1	34.8	24.4	23.3	2.6
Percent variance	85.66	6.65	4.82	1.68	0.49	0.34	0.32	0.04
Cumulative percent variance	85.66	92.31	97.13	98.81	99.3	99.64	99.96	100.0

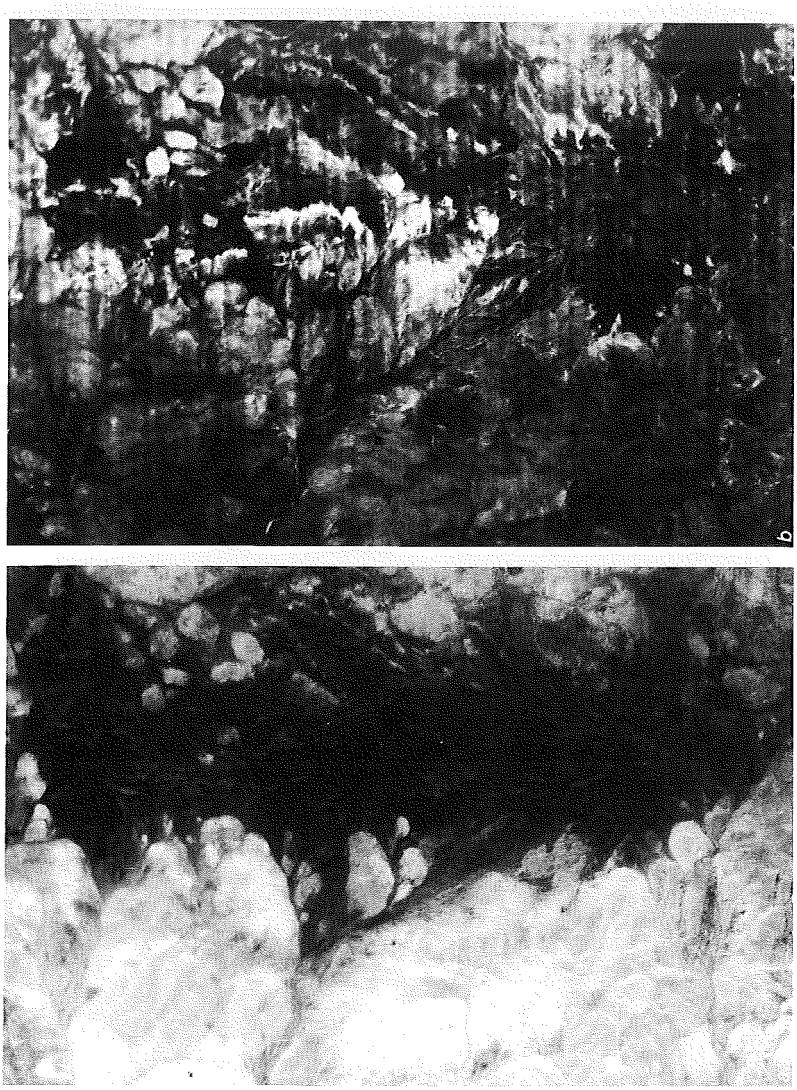


FIGURE 4.19—*Principal components 1 (a) and 2 (b) images computed from the multispectral image in figure 4.18.*

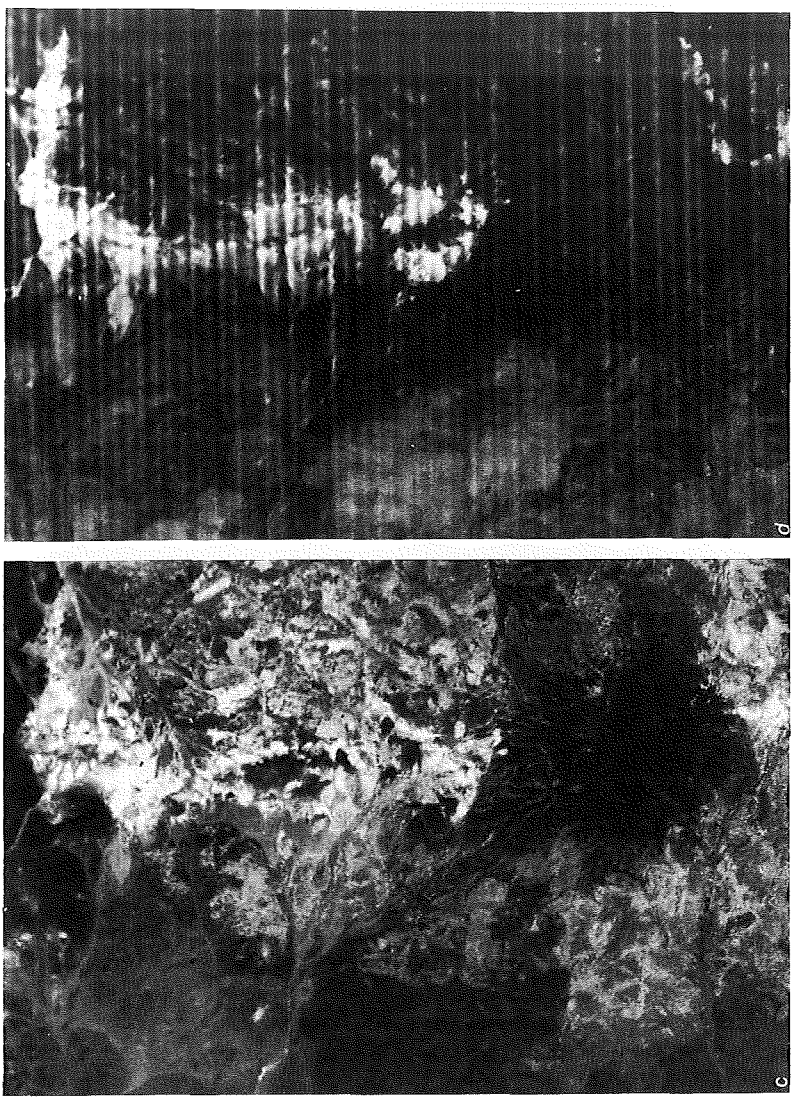


FIGURE 4.19—*Continued. Principal components 3 (c) and 4 (d) computed from the multispectral image in figure 4.18.*

FIGURE 4.19—*Continued. Principal components 5 (e) and 6 (f) computed from the multispectral image in figure 4.18.*

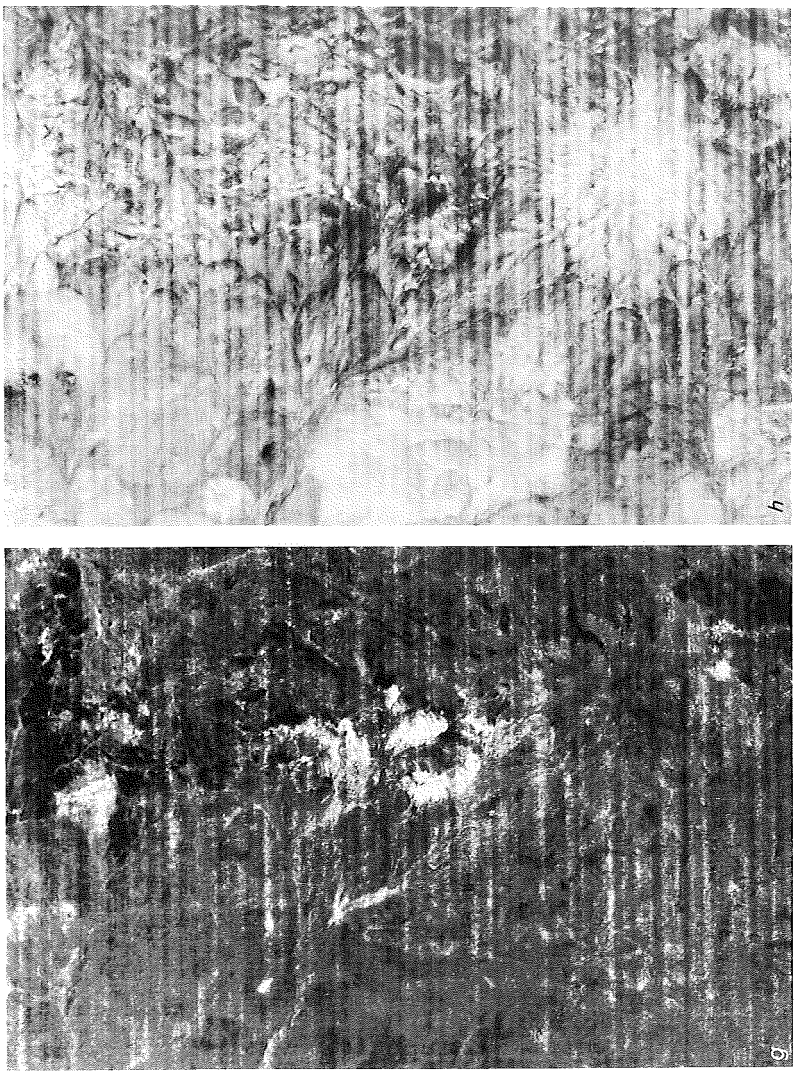


FIGURE 4.19—*Continued. Principal components 7 (g) and 8 (h) computed from the multispectral image in figure 4.18.*

FIGURE 4.20a—*False-color display of principal component images. Original channels 3 (red), 6 (green), and 8 (blue).*

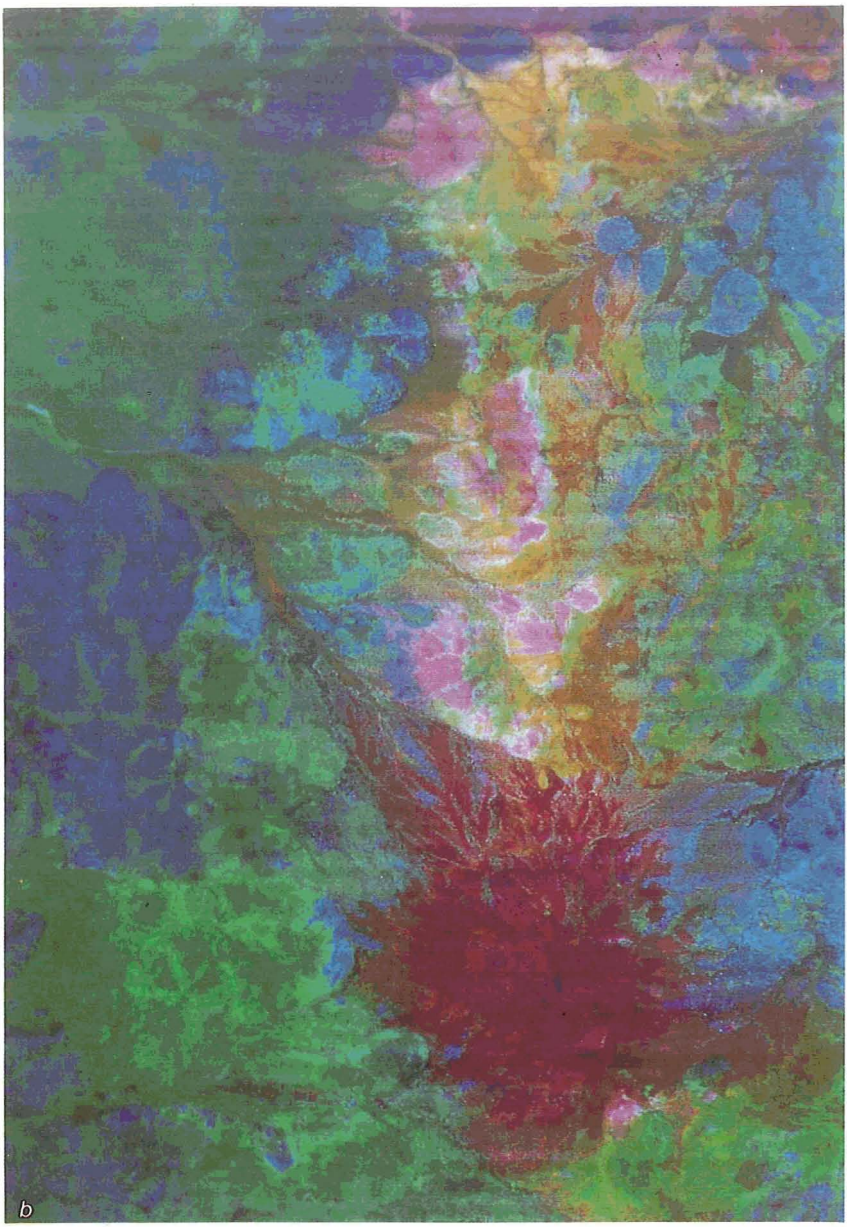


FIGURE 4.20*b*—*False-color display of principal component images. Principal components 1 (red), 2 (green), and 3 (blue).*

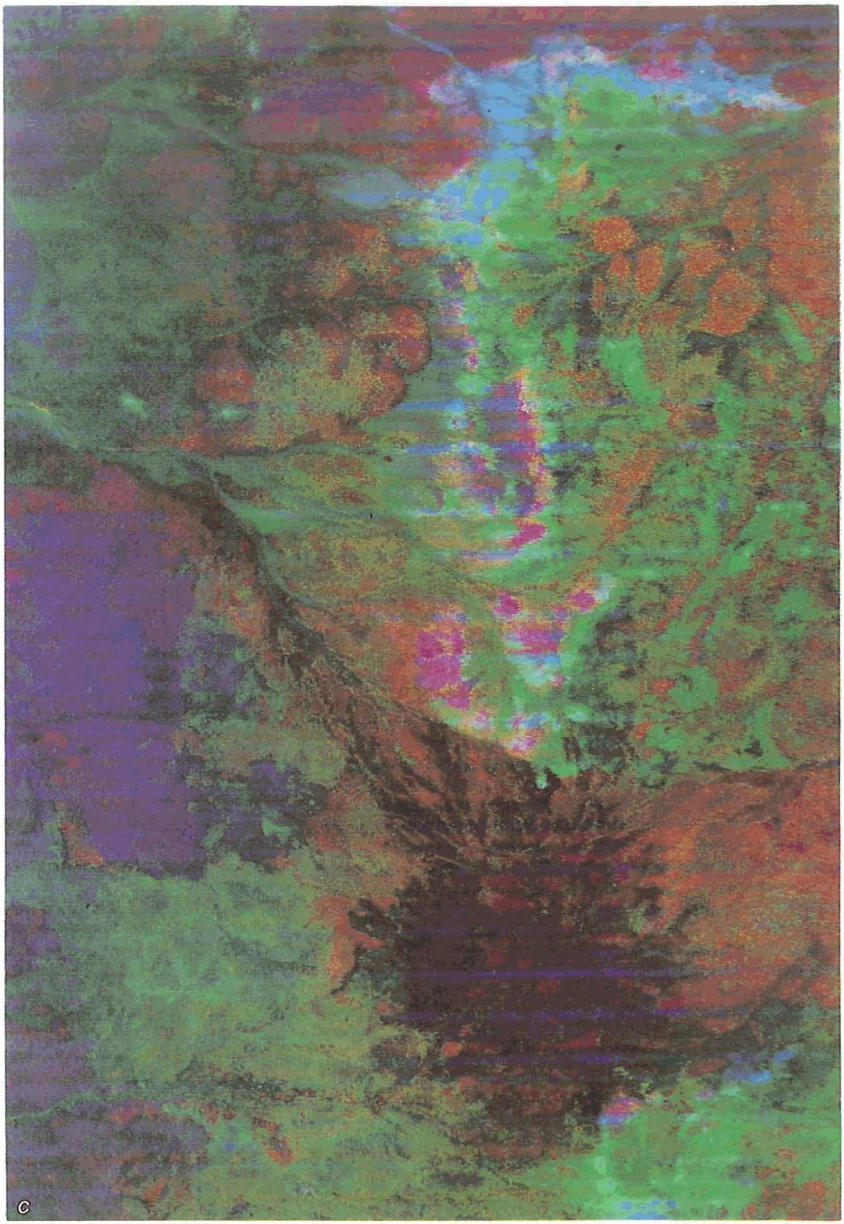


FIGURE 4.20c—*False-color display of principal component images. Principal components 2 (red), 3 (green), and 4 (blue).*

REFERENCES

[1] Schwartz, A. A.: New Techniques for Digital Image Enhancement. Proceedings of Caltech/JPL Conference on Image Processing Technology, Data Sources and Software for Commercial and Scientific Applications, California Institute of Technology, Pasadena, Calif., Nov. 1976, pp. 5-1–5-10.

[2] Levi, L.: On Image Evaluation and Enhancement, Opt. Acta, vol. 17, 1970, pp. 59–76.

[3] Campbell, F. W.: The Human Eye as an Optical Filter, Proc. IEEE, vol. 56, 1968, pp. 1009–1014.

[4] Fink, W.: Image Coloration as an Interpretation Aid. Proceedings of OSA/SPIE Meeting on Image Processing, Asilomar, Calif., vol. 74, 1976.

[5] Andrews, H. C.; Tescher, A. G.; and Kruger, R. P.: Image Processing by Digital Computer, IEEE Spectrum, vol. 9, no. 7, 1972, pp. 20–32.

[6] Nathan, R.: Picture Enhancement for the Moon, Mars and Man, *in* Cheng, G. C., et al., eds.: Pictorial Pattern Recognition. Thompson, Washington, D.C., 1968, pp. 239–266.

[7] Selzer, R. H.: Improving Biomedical Image Quality with Computers. NASA JPL TR 32-1336, 1968.

[8] Huang, T. S.: Image Enhancement: A Review, Opto-Electronics, vol. 1, 1969, pp. 49–59.

[9] Rosenfeld, A.; and Kak, A. C.: Digital Picture Processing. Academic Press, New York, 1976.

[10] Hummel, R. A.: Histogram Modification, Computer Graphics and Image Processing, vol. 4, 1975, p. 209, and vol. 6, 1977, p. 184.

[11] O'Handley, D. A.; and Green, W. B.: Recent Developments in Digital Image Processing at the Image Processing Laboratory at the Jet Propulsion Laboratory, Proc. IEEE, vol. 60, 1972, pp. 821–828.

[12] Prewitt, M. S.: Object Enhancement and Extraction, *in* Lipkin, B. S.; and Rosenfeld, A.: Picture Processing and Psychopictorics. Academic Press, New York and London, 1970, pp. 75–149.

[13] Podwysoki, M. H.; Moik, J. G.; and Shoup, W. C.: Quantification of Geologic Lineaments by Manual and Machine Processing Techniques. NASA Goddard Space Flight Center, X-923-75-183, July 1975.

[14] Seidman, J.: Some Practical Applications of Digital Filtering in Image Processing. Proceedings of Computer Image Processing and Recognition, University of Missouri, Columbia, Mo., Aug. 1972.

[15] Stockham, T. S.: Image Processing in the Context of a Visual Model, Proc. IEEE, vol. 60, 1972, pp. 828–842.

[16] Billingsley, F. C.; Goetz, A. F. H.; and Lindsley, J. N.: Color Differentiation by Computer Image Processing, Photogr. Sci. Eng., vol. 17, 1970, pp. 28–35.

[17] Billmeyer, F. W.; and Saltzmann, M.: Principles of Color Technology. Interscience, New York, 1966.

[18] Sheppard, J. J.; Stratton, R. H.; and Gazley, C. G.: Pseudocolor as a Means of Image Enhancement, Am. J. Optom., vol. 46, 1969, pp. 735–754.

[19] Billingsley, F. C.: Some Digital Techniques for Enhancing ERTS Imagery. American Society of Photogrammetry, Sioux Falls Remote Sensing Symposium, Sioux Falls, N. Dak., Oct. 1973.

[20] Goetz, A. F. H.; Billingsley, F. C.; Gillespie, A. R.; Abrams, M. J.; and Squires, R. L.: Application of ERTS Image and Image Processing to Regional Geologic Problems and Geologic Mapping in Northern Arizona. NASA/JPL TR 32-1597, May 1975.

[21] Stouffer, M. L.; and McKinney, R. L.: Landsat Image Differencing as an Automated Land Cover Change Detection Technique. Computer Sciences Corp., TM-78/6215, Aug. 1972.

[22] Readv, P. J.; and Wintz, P. A.; Information Extraction, SNR Improvement, and Data Compression in Multispectral Imagery, IEEE Trans. Commun., vol. COM-21, 1973, pp. 1123–1131.

[23] Zaitzeff, J. M.; Wilson, C. L.; and Ebert. D. H.: MSDS: An Experimental 24-Channel Multispectral Scanner System, Bendix Technical Journal, vol. 3, no. 2, 1970, pp. 20–32.

Page intentionally left blank

5. Image Registration

5.1 Introduction

In many image processing applications it is necessary to compare and analyze images of the same scene obtained from different sensors at the same time, or taken by one or several sensors at different times. Such applications include multispectral, multitemporal, and multisensor statistical pattern recognition, change detection, and map matching for navigation [1]. Multiple measurements from each resolution element provide a means for detecting time varying properties and for improving the accuracy of recognition. A collection of images of the same scene is called a multi-image.

An implicit assumption in the analysis of multi-images is that the component images are registered, i.e., that a measurement vector in a multi-image is derived from a common ground resolution element. However, multiple images of the same scene are generally not registered, but are spatially distorted relative to one another. Misregistration results from the inability of sensing systems to produce congruent measurements because of design characteristics or because accurate spatial alinement of sensors at different times is impossible. Relative translational and rotational shifts and scale differences, as well as geometric and intensity distortions, can all combine to produce misregistration.

Therefore, image registration is required before analysis procedures can access contextually coincident resolution elements in each component of a multi-image by one unique coordinate pair. (See fig. 5.1.) Image registration is the procedure that generates a set of spatially alined, or matches, images of a scene. The registration procedure consists of two steps, (1) the determination of matching context points in multiple images, and (2) the geometric transformation of the image so that the registration of each context point is achieved. In this chapter only techniques for the first step are described. The geometric transformation of images is discussed in section 3.3.

Practically, two methods of image registration may be distinguished. For relative registration, one component of a multi-image is selected as the reference to which the other component images are to be registered. For absolute registration, a control grid (e.g., given by a cartographic projection) is defined, and all component images are registered to this reference.

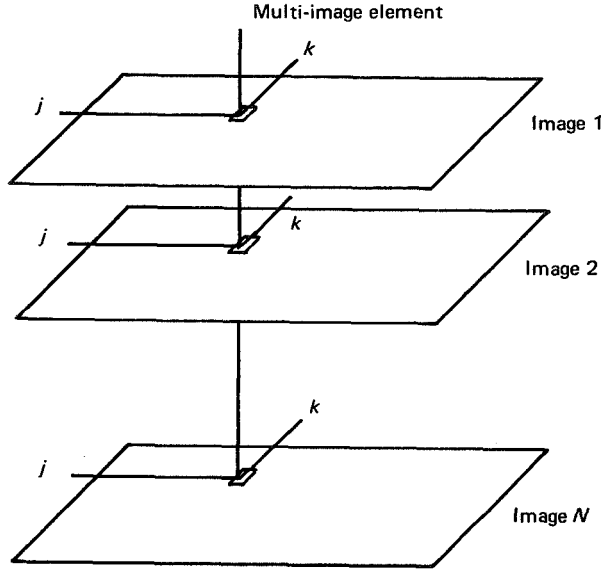


FIGURE 5.1—*Multi-image registration.*

If the geometric distortions are exactly the same for all images to be registered, the alinement is accomplished by determining the relative translation between the images. In situations in which the relative spatial distortions between the images are small, it can be assumed that the spatial differences are negligible for small regions. Registration is then accomplished by determining the relative translation of subimages and applying a geometric transformation based on the displacement of the subimages.

The determination of corresponding points in the component images is a problem of template matching. Subareas of the reference that contain invariant features are extracted and referred to as templates. Corresponding subareas of the images to be registered are selected as search areas (fig. 5.2).

A search area $\mathbf{S}$ is a matrix of $J \times K$ picture elements. A template $\mathbf{T}$ is a matrix of $M \times N$ elements. It is assumed that a search area is larger than a template ($J > M, K > N$) and that enough a priori information is available about the displacement between the images to permit selection of the location and size of templates and search areas such that, at registration, a template is completely contained in its search area (fig. 5.3).

The problem is to determine at which location (j^*, k^*) the template matches the corresponding search area. The existence of a matching location can be assumed, but because of geometric and intensity distor-

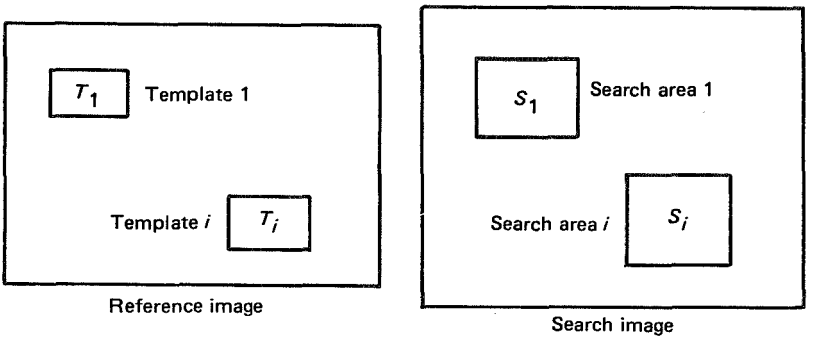


FIGURE 5.2—*Selection of templates and search areas.*

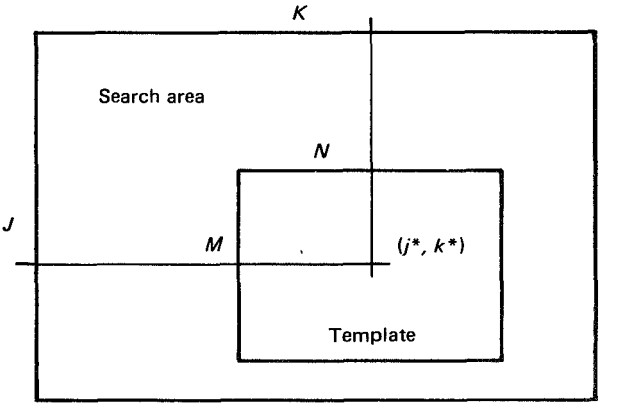


FIGURE 5.3—*Template and search area.*

tions, real changes in the scene, and noise, there is no way to make certain that a correct match has been achieved. At most, the probability that the images are in a certain geometrical relationship to each other can be determined from the available data. The optimum registration algorithm would produce a set of a posteriori probabilities describing each possible relationship. Once these probabilities are determined, a statistical decision rule can be defined by the requirement that some measure of the cost of a decision be minimum. However, because the characteristics of the distortions and noise that define the relationship between a template and its mapping in the search area are in general unknown, the computation of the required probability density distributions is practically impossible.

Therefore, approximations in the form of maximizing a similarity measure are used. The decision regarding the location of a match is made by searching for the maximum of the similarity measure and comparing it with a predetermined threshold. Generally, there is no theoretically derived evaluation of the error performance of a registration technique before its actual application [2].

5.2 Matching by Crosscorrelation

The similarity between two images f and g over a region S can be measured in several ways [3]. Commonly used similarity or distance measures are the quadratic difference

$$d_P = \sum_{j=1}^{M} \sum_{k=1}^{N} [f(j,k) - g(j,k)]^2 \tag{5.1}$$

and the absolute difference

$$d_A = \sum_{j=1}^{M} \sum_{k=1}^{N} |f(j,k) - g(j,k)| \tag{5.2}$$

Equation (5.1) can be expressed by using the Cauchy-Schwartz inequality

$$\sum_j \sum_k f(j,k)g(j,k) \leq \left(\sum_j \sum_k f(j,k)^2 \sum_j \sum_k g(j,k)^2 \right)^{1/2} \tag{5.3}$$

with the equality holding if and only if $g(j,k) = cf(j,k)$ for all j and k and with c constant. Thus, when the following quantities are given:

$$\sum_j \sum_k f(j,k)^2 \quad \text{and} \quad \sum_j \sum_k g(j,k)^2$$

d_P is given by

$$d_P = \sum_{j=1}^{M} \sum_{k=1}^{N} f(j,k)g(j,k) \tag{5.4}$$

and is a measure of the degree of match between f and g.

For template matching it is assumed that f represents a template T and that g represents a search area S. The problem is to find parts of g that match f. For this search f is shifted into all possible positions relative to g, and for each position (m, n) the distances d_P or d_A are computed. Thus, equation (5.3) becomes

$$\sum_j \sum_k f(j,k)g(j+m,k+n) \leq \left[\sum_j \sum_k f(j,k)^2 \sum_j \sum_k g(j+m,k+n)^2 \right]^{1/2} \tag{5.5}$$

The left-hand side of equation (5.5) is the crosscorrelation between f and g. The right-hand side depends on m, and, therefore, the normalized crosscorrelation is used as a measure of match:

$$R(m, n) = \frac{\sum_j \sum_k f(j, k) g(j+m, k+n)}{\left[\sum_j \sum_k f(j, k)^2 \sum_j \sum_k g(j+m, k+n)^2\right]^{1/2}} \tag{5.6}$$

To compensate for differences in illumination, f and g may be normalized by subtracting their average gray values $\bar{f}$ and $\bar{g}$. This step yields

$$\bar{R}(m, n) = \frac{\sum_j \sum_k f(j, k) g(j+m, k+n) - \bar{f}\bar{g}(m, n)}{\sigma_f \sigma_{g(m, n)}} \tag{5.7}$$

where σ_f is the standard deviation of the gray values of f and $\sigma_{g(m, n)}$ is the standard deviation of the gray values of g in an area of the size of f at location (m, n), respectively.

The variable R takes on its maximum value for displacements (m^*, n^*) at which $g = cf$, i.e., for a perfect correlation between f and g. Thus, template matching involves the computation of the similarity measures d_P for each possible displacement, and search for the displacement (m^*, n^*) at which d_P is maximum.

For the absolute difference similarity measure,

$$S(m, n) = \sum_j \sum_k |f(j, k) - g(j+m, k+n)| \tag{5.8}$$

is computed, and the displacement (m^*, n^*) at which $S(m, n)$ is minimum is the location of best match. This metric is used in the sequential similarity detection algorithm [4].

The correlation measures given in equations (5.6) or (5.8) determine only translational differences. It is assumed that f and g have the same scale and rotational alinement. This alinement does not occur in general, and scaling and rotational transformations must be carried out in addition to translation if there are severe scale and angular differences between f and g. Because of the presence of noise and distortions between f and g, there is only a certain probability that the extrema of equation (5.6) or (5.8) actually define the correct match point.

5.3 Registration Errors

Geometric distortions and systematic intensity changes are the main errors that can degrade the correlation performance.

5.3.1 Geometric Distortions

Any geometric distortion of the search image coordinates relative to the reference image coordinates degrades the performance of the registration process. The most important types of geometric distortions are scale, rotation, skew, scan nonlinearity, and perspective errors. Geometric error sources are discussed in section 2.4.1.

Scale errors are primarily caused by altitude changes. The template elements are either somewhat larger or smaller than the search-area elements. Consequently, elements of the template, when overlaid on the search area, encompass both matching and nonmatching elements, and the amount of nonmatching overlap increases outward from the center.

Rotation errors can be caused by attitude or heading changes. If the template is centered but rotated relative to the search area, the correlation algorithm compares a single template element with a combination of fractions of both matching and nonmatching search-area elements. The amount of overlap with nonmatching elements increases outward from the center of the template.

Skew errors in satellite scanner images are caused by the Earth's rotation between successive mirror sweeps. Again the correlation algorithm compares a single template element with both matching and nonmatching search-area elements.

Scan nonlinearity errors are caused by the nonlinear motion of the scanner, resulting in a nonlinear scale change in the scan line direction. Scan length variations, caused by changes in the period of the oscillating scan mirror, are often corrected in ground processing, e.g., synthetic pixels in Landsat MSS. These pixels must be removed before registration. Perspective errors occur when the reference and search images were taken from different positions. The effect is similar to a linearly varying scale-factor error. The geometric distortions between reference and search images are shown in figure 5.4.

5.3.2 Systematic Intensity Errors

Systematic intensity errors include all changes in the intensity of the search image, relative to the reference image, that cannot be attributed to sensor noise.

The overall signal level of the search image relative to the reference image can be altered by changes in scene illumination (e.g., day to night or sunny to overcast) or by changes in sensor gain settings. Changes in the optical properties of the atmosphere can also change the overall signal level, the contrast perceived by the sensor, or both. Shadows due to clouds or changes in Sun angle cause blocks of search image elements to be totally dissimilar to the corresponding reference image elements. The reflectivity of certain portions of a scene can change as a result of

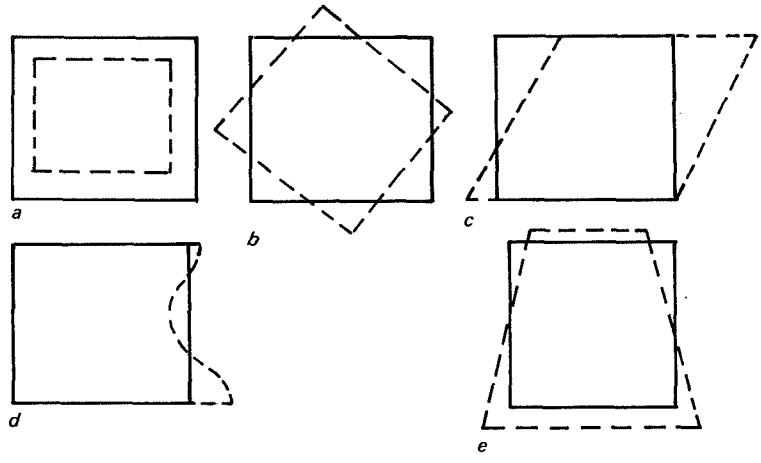


FIGURE 5.4—*Geometric distortions between reference and search images. Solid figures are the reference images; dashed figures are the search images. (a) Scale. (b) Rotation. (c) Skew. (d) Scan nonlinearity and scan length variations. (e) Perspective.*

physical changes on the ground, such as snowfall or flooding; as a result of differences in moisture content or seasonal changes in foliage and vegetation; or, to a lesser degree, simply as a result of differences in the direction of the illumination by either active sensors or the Sun changing orientation at different times of day. Finally, the search image can be different from the reference image owing to actual changes in the reference scene (e.g., new manmade objects).

These systematic errors generally do not significantly increase the width of the correlation function, but they reduce the differential between the in- and out-of-register values, and thereby increase the possibility for false correlations. Uniform intensity or gain changes by a factor c do not affect the performance of algorithms using equation 5.6 because this measure of match attains its maximum whenever the search area and template values are proportional to each other. The similarity measure given in equation (5.8) attains its minimum when the template and search area are equal to each other. Thus, changes in the overall signal level can severely influence the performance of algorithms using this equation.

The question of whether equation (5.6) or (5.8) should be used depends on the image statistics that have been ignored in the definition of the similarity measures. Experiments have shown that at low signal-to-noise ratios (SNRs)—SNR ≤ 1—algorithms using the normalized correlation coefficient perform better [5]. At high SNRs (SNR ≥ 3), absolute-difference algorithms are better. However, in practical applications, these

high SNRs are seldom realized. When $1 < \text{SNR} < 3$, the choice of algorithm is not critical. (For the computation of the SNR, it is assumed that the template f has variance σ_f^2 and that the search image g is the template corrupted by noise n, such that $g=f+n$, where n has zero mean and variance σ_n^2. Then $\text{SNR}=\sigma_f^2/\sigma_n^2$).

5.3.3 Preprocessing for Image Registration

If severe geometric distortions exist between the reference and search images, crosscorrelation will not yield useful results unless the distortions are first estimated and removed. If the reference and search image areas containing the same object differ substantially in average gray level, their gray-level distribution should be normalized. After the geometric and contrast correction, the translational misregistration is estimated for each subarea by determining the location of the peak of the correlation surface. This surface is computed by crosscorrelating a template, obtained from the reference image, with the corresponding search area. The peak of this correlation surface is assumed to be the point of correct superposition of the template on the search area. In general, the correlation surface computed by equation (5.6) may be rather broad, making detection of the peak difficult.

5.4 Statistical Correlation

The detection of the correlation peak may be facilitated by including the statistical properties of the reference image into the correlation measure [6, 7]. The objective is to filter the template f so that the correlation measure corresponding to the correct superposition of f on the search area g is maximally discriminable from the correlation results of all other positions of f or g. This idea can be implemented as maximizing the ratio of the correlation measure at the matching location to the variance of its values taken over all other locations [8]. The statistical properties of the reference image are assumed to be characterized by the spatial covariance matrix $\mathbf{C}$.

The linear space invariant filter h is chosen such that the ratio

$$Z=\frac{R_s^2(m^*, n^*)}{\text{var}\,[R_s(m, n)]} \tag{5.9}$$

is maximized. The statistical correlation measure R_s is defined as

$$R_s(m, n)=\frac{\sum_j \sum_k s(j, k)g(j+m, k+n)}{\left[\sum_j \sum_k s^2(j, k)\right]^{1/2}\left[\sum_j \sum_k g^2(j+m, k+n)\right]^{1/2}} \tag{5.10}$$

where the new template s is obtained by convolving f with the filter h; i.e., $s=f * h$. Determination of the optimal registration filter requires

computation of the covariance matrix and its inverse. This operation is numerically difficult, because for an M by N template, $\mathbf{C}$ will be of dimension MN by MN. Furthermore, a large set of data is required for estimating the covariance matrix. By making simplifying assumptions about the statistical nature of the images to be registered, it is possible to reduce the computational problems significantly. If it is assumed that the statistical properties of the reference image are modeled by an isotropic exponential covariance matrix [6, 7] $\mathbf{C}$, given by

$$\mathbf{C}=(\rho^{|i-j|}) \tag{5.11}$$

where ρ is the average adjacent element correlation coefficient, then the registration filter is given by

$$[h]=\begin{pmatrix} \rho^2 & -\rho(1+\rho^2) & \rho^2 \\ -\rho(1+\rho^2) & (1+\rho^2)^2 & -\rho(1+\rho^2) \\ \rho^2 & -\rho(1+\rho^2) & \rho^2 \end{pmatrix} \tag{5.12}$$

If the images are completely spatially uncorrelated ($\rho=0$), the filter is given by

$$[h]=\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{pmatrix} \tag{5.13}$$

and the template is directly taken from the reference image. For completely correlated images ($\rho=1$), the filter becomes

$$[h]=\begin{pmatrix} 1 & -2 & 1 \\ -2 & 4 & -2 \\ 1 & -2 & 1 \end{pmatrix} \tag{5.14}$$

This equation is the discrete approximation to the mixed fourth partial derivative, obtained by convolving the discrete approximations to the second partial derivatives along each coordinate axis:

$$[h]=[h_1]*[h_2]\begin{pmatrix} 0 & 0 & 0 \\ 1 & -2 & 1 \\ 0 & 0 & 0 \end{pmatrix} * \begin{pmatrix} 0 & 1 & 0 \\ 0 & -2 & 0 \\ 0 & 1 & 0 \end{pmatrix} \tag{5.15}$$

Thus, when the images are highly correlated, the correlation concentrates on the edge comparison between the reference and search images.

The use of derivatives of f as registration filters can also be heuristically justified. If the features of interest in the template and search area are characterized by shape rather than by contrast, the edge images of f and g can be correlated. Correlating edge images tends to yield sharper matches than does correlation of gray-value images. Edge images also permit the

use of multispectral shape information in the compact form of a single composite edge image [9–11]. An evaluation and comparison of similarity measures is given in [12] and [13].

5.5 Computation of the Correlation Function

In general, the correlation function must be computed for all possible translations of the template within the search area to determine its maximum value and obtain an estimate for the misregistration. The number L of these translations is given by $L=(J-M+1)(K-N+1)$. Finding the maximum value of the correlation function requires MN multiplications to be performed for each of L relative shifts. For larger f and g the computation time can be reduced by applying the fast Fourier transform (FFT) algorithm [14]. By the convolution theorem, equation (2.46), of Fourier analysis, crosscorrelating a template f with a search area g is equivalent to pointwise multiplying the Fourier transforms F^* and G and then taking the inverse transform:

$$R(m, n) = \mathcal{F}^{-1}[F^*(u, v)\ G(u, v)] \tag{5.16}$$

where $\mathcal{F}$ denotes the Fourier transform operator, and u and v are spatial frequencies. Correlation functions obtained with the discrete Fourier transform (DCT) are cyclic, because the transform assumes the pictures to be periodic functions. (See sec. 2.6.1.1.)

Thus, cyclic convolutions have values even for shifts such that the template is no longer entirely inside the picture. The Fourier transform matrices to be multiplied pointwise must be of the same size. Therefore, the template is extended by zeroes to the size of the search area. The valid part of the computed correlation function is rearranged for determination of the correlation maximum and for display. (See sec. 2.6.3.) An estimate of the location of the correlation peak to subpixel accuracy may be obtained by fitting a bivariate polynomial to $R(m, n)$ and computing its maximum.

REFERENCES

[1] Littestrand, R. L.: Techniques for Change Detection, IEEE Trans. Comput., vol. C-21, 1972, pp. 654–659.

[2] Pinsin, L. J.; Boland, J. S.; and Malcolm, W. W.: Statistical Analysis for a Binary Image Correlator in the Absence of Geometric Distortion, Opt. Eng., vol. 6, 1978, pp. 635–639.

[3] Rosenfeld, A.; and Kak, A. C.: Digital Picture Processing. Academic Press, New York, 1976.

[4] Barnea, D. I.; and Silverman, H. F.: A Class of Algorithms for Fast Digital Image Registration, IEEE Trans. Comput., vol. C-21, 1972, pp. 179–186.

[5] Bailey, H. H., et al.: Image Correlation: Part 1, Simulation and Analysis, Rand Corp. Report R-2057/1-PR, 1976.

[6] Arcese, A.; Mengert, P. H.; and Trombini, E. W.: Image Detection through Bipolar Correlation, IEEE Trans. Info. Theory, vol. IT-16, 1970, pp. 534–541.
[7] Pratt, W. K.: Correlation Techniques of Image Registration, IEEE Trans. on Aerosp. and Electron. Syst., vol. AES-10, 1974, pp. 353–358.
[8] Emmert, R. A.; and McGillem, C. D.: Conjugate Point Determination for Multitemporal Data Overlay. LARS Information Note 111872, Purdue University, Lafayette, Ind., 1973.
[9] Nack, M. L.: Temporal Registration of Multispectral Digital Satellite Images Using Their Edge Images. AAS/AIAA Astrodynamics Specialist Conference, Nassau, Bahamas, July 1975.
[10] Nack, M. L.: Rectification and Registration of Digital Images and the Effect of Cloud Detection. Proceedings of Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Lafayette, Ind., 1977, pp. 12–23.
[11] Jayroe, R. R.; Andrus, J. F.; and Campbell, C. W.: Digital Image Registration Method Based upon Binary Boundary Maps. NASA TND-7607, Washington, D.C., Mar. 1974.
[12] Svedlow, M.; McGillem, C. D.; and Anuta, P. E.: Experimental Examination of Similarity Measures and Preprocessing Methods Used for Image Registration. Proceedings of Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Lafayette, Ind., 1976, pp. 4A-9–4A-13.
[13] Kaneko, T.: Evaluation of Landsat Image Registration Accuracy, Photogr. Eng. and Remote Sensing, vol. 42, 1976, pp. 1285–1299.
[14] Anuta, P. E.: Spatial Registration of Multispectral and Multitemporal Digital Imagery Using Fast Fourier Transform Techniques, IEEE Trans. Geosci. Electron, vol. 8, 1970, pp. 353–368.

Page intentionally left blank

6. Image Overlaying and Mosaicking

6.1 Introduction

The generation of image overlays and mosaics is a key requirement for image analysis. Overlaying is the spatial superposition of images taken at different wavelengths, at different times, or by different sensors such that congruent measurements for each raster element are obtained. Multiple congruent spatial distributions were defined as a multi-image in section 2.1. Congruent measurements are required for many image analysis applications such as multidimensional classification, change detection, and modeling. For example, overlaying multispectral and multitemporal measurements with data from other sensors and with ancillary information (e.g., terrain elevation and slope) offers a means of improving the recognition accuracy in classification.

Change detection involves the comparison of two images of the same scene taken at different times. The problem is to detect the amount of change of image properties rather than their absolute magnitude. A possible application is detection of the change in the width of a river, in the size of a lake after a storm, or in land use or urban patterns. An overlay of a temporal sequence may be used for trend analysis. Overlaying two images taken from different positions in space results in a pair of stereoscopic images that enable calculations of the third spatial coordinate of every image point. For example, cloud-height analysis can be performed with images taken by two geosynchronous satellites. Terrain elevation images may be created from stereoscopic image pairs.

Map overlays of images and of image analysis results can be used to update maps and to produce new maps. Overlays of spatially congruent data from various sensors are required as input to climatic, environmental, and land-use models.

Mosaicking is the combination of several image frames into photomosaics covering a specified area. Such mosaics can be used for mapmaking. In general, the frame size is given by the field of view of the sensor, and the frame location is determined by spacecraft operation. Thus, an area of interest may be partly covered by several frames, each with its particular geometric distortions and radiometric degradations.

6.2 Techniques for Generation of Overlays and Mosaics

The generation of image overlays and mosaics requires similar image processing techniques. The image frames are usually taken with different

attitudes and positions of the sensors, at different times and seasons, and under different atmospheric conditions. Varying geometric distortions prevent accurate overlay of corresponding frames. Scale and shape differences in adjacent images may be so severe that a set of frames cannot be mosaicked without misalinement at boundaries.

Radiometric differences in adjacent frames caused by Sun-angle-dependent shadows, seasonal changes of fields, forests, water bodies, and different atmospheric conditions may produce artificial edges in mosaics. Clouds and noise in the border area of one frame can also produce discontinuities at the seams between images. Therefore, geometric and radiometric corrections and transformation to a common reference are required for overlaying and mosaicking images. Two basic approaches are available for generation of overlays and mosaics. For overlays, one image may be selected as reference and the other frames are then registered to this reference. Techniques for geometric transformation and image registration are discussed in chapters 3 and 5. The second approach is to select a cartographic projection and to register all images to this common map reference. Map projections will be discussed in section 6.3.

Similarly, mosaics may be produced by selecting one frame as reference and registering adjacent frames to the reference. This operation requires a sufficiently large area of overlap between adjacent frames, and only limited geometric accuracy may be achieved. This approach is thus limited to the generation of mosaics consisting of only a few frames. The second approach is to choose a cartographic projection as a reference grid and to transform all frames to it. Map projections are continuous representations of a surface. Therefore, a set of frames transformed to the same projection will mosaic perfectly.

Map projections are a basis for a standard representation of discrete spatially distributed measurements, such as digital remote sensing images and point measurements and ground truth data. To relate these data, a common framework in the form of a well-defined coordinate system is required. The location of each measurement on the surface of the earth is uniquely defined by the geographic or geodetic coordinates (longitude λ, latitude φ) and the elevation z above sea level. A map projection defines the transformation of data locations from geographic to plane coordinates and provides the common framework for analysis, graphical display, and building of a data base.

6.3 Map Projections

A map projection is the representation of a curved surface in a plane. For the Earth, the curved reference or datum surface is assumed to be an ellipsoid or a sphere. Projection surfaces are planes, cylinders, and cones, and the latter two types are developable into planes [1]. The transformation from the datum to the projection surface is defined by a set of

mathematical expressions describing the relationship between latitudes and longitudes in the datum surface and the coordinates in the projection plane, which is dependent on the type of projection. Any projection of a curved surface onto a plane involves distortions of distances, shapes, or areas [2]. Map projection criteria that preserve distance, shape, and area are mutually exclusive. Therefore, there is no ideal projection, but only a best planar representation for a given purpose.

6.3.1 Classes of Map Projections

Map projections are divided into classes according to the following criteria:

1. The nature of the geometric properties of the projection surface
2. The contact of the projection surface with the datum surface
3. The position of the projection surface with relation to the datum surface

These classes are not mutually exclusive. Criterion 1 leads to planar, cylindrical, and conical projections, each representing one of the basic projection surfaces: plane, cylinder, and cone. The simplest of these projection surfaces is the plane, which when tangent to the datum surface, would have a single point of contact, this point also being the center of the area of minimum distortion. The cone and the cylinder, which are both developable into a plane, increase the extent of contact and, consequently, the area of minimum distortion. (See fig. 6.1.)

Criterion 2 yields three groups of projections, representing three types of contact between the datum and projection surface: tangent, secant, and polysuperficial. Tangency between the datum and projection surfaces results in a point contact if the projection surface is a plane and a line contact if the projection surface is either a cone or a cylinder. In the secant case a line of contact is obtained when the projection surface is a plane, and two lines of contact are obtained when the projection surface is either a cone or a cylinder. These principles are illustrated in figure 6.2 for the plane and cone.

A further increase of contact between the datum and projection surfaces, and thus a reduction of the distortion, is achieved by a series of successive projection surfaces. A series of planes would produce a polyhedric (multiple-plane) projection; a series of cones, a polyconic; and a series of cylinders, a polycylindrical projection.

Criterion 3 leads to subdivision into three groups, representing the three basic positions of the projection surface relative to the datum surface: normal, transverse, and oblique. If the purpose of the projection is to represent a limited area of the datum surface, it is advantageous to achieve the minimum of distortion for that particular area. Minimizing distortion is possible by varying the attitude of the projection surface. If

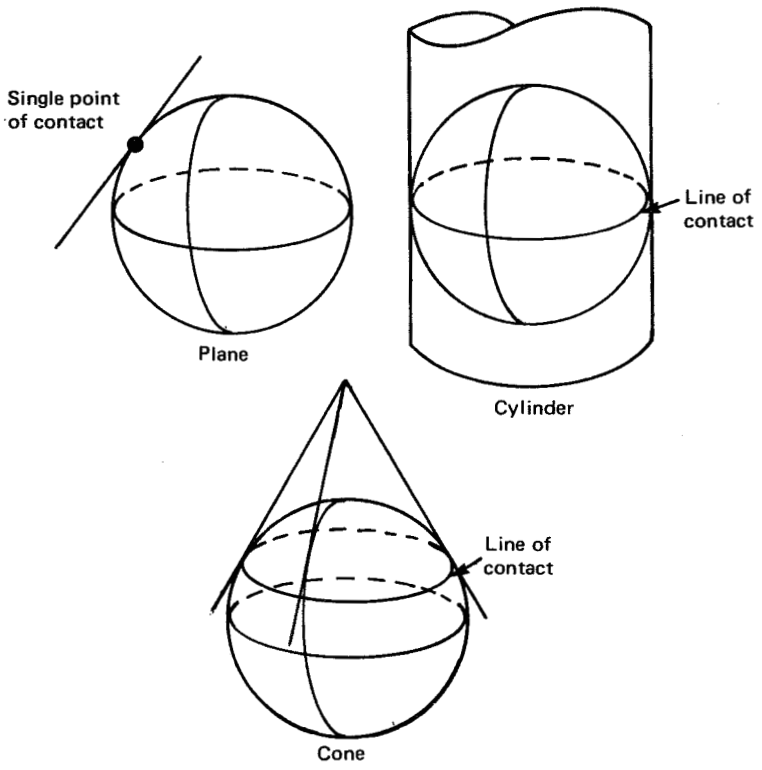


FIGURE 6.1—*Projection surfaces (after Richardus and Adler [1]).*

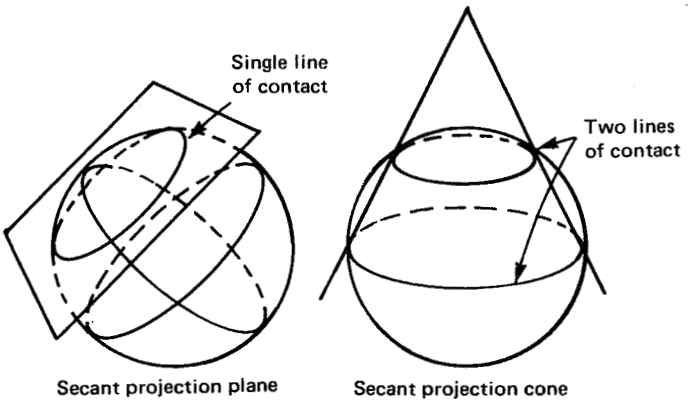


FIGURE 6.2—*Increase of contact between projection and datum surface.*

the axis of symmetry of the projection surface coincides with the rotational axis of the ellipsoid or the sphere, the normal case is obtained. With the axis of symmetry perpendicular to the axis of rotation, the transverse projection is obtained. Any other attitudes of the axis of symmetry result in oblique projections. (See fig. 6.3.)

Projections may also be characterized according to the cartographic properties equidistance, conformality, and equivalency. These properties are mutually exclusive. Equidistance is the correct representation, on the projection surface, of the distance between two points of the datum surface. This property is not a general one, and it is limited to certain specified points. Conformality means the correct representation of the shape or form of objects. This property may be limited to small areas. Equivalency is the correct representation of areas on the projection surface at the expense of shape distortions.

6.3.2 Coordinate Systems

Coordinate systems are required to relate points on the datum and projection surfaces. The datum surface of the Earth is usually an ellipsoid or sphere with the coordinates expressed as longitude λ, counted positive from a reference meridian, and latitude φ, counted positive from the equator (fig. 6.4).

The coordinate system in the projection plane is a rectangular Cartesian system (x, y) with the positive y-axis pointing north (sometimes referred to as northing), and the positive x-axis pointing east (easting). The coordinate systems may be graphically represented by regularly spaced grids of longitudes and latitudes, or northings and eastings. A map projection is the transformation of grids from the curved surface to the projection plane. The origin is usually the central point of the projected area. With cylindrical or conical projections, this central point may be located on the tangent parallel or meridian.

The relationship between the projection plane and the ellipsoidal or spherical coordinate system is given by

$$(x, y) = \mathbf{T}_p(\lambda, \varphi) \tag{6.1}$$

where $\mathbf{T}_p$ is a vector function determined by the type of projection.

6.3.3 Perspective Projections

Perspective projections are projections onto a plane that is perpendicular to a line through the center of the sphere or perpendicular to an ellipsoidal normal. Thus, the projection axis is perpendicular to both the datum and the projection surfaces, and the projection center lies at the intersection of the projection axis and the projection plane. A point on the projection axis serves as the perspective point, and straight lines from the perspective

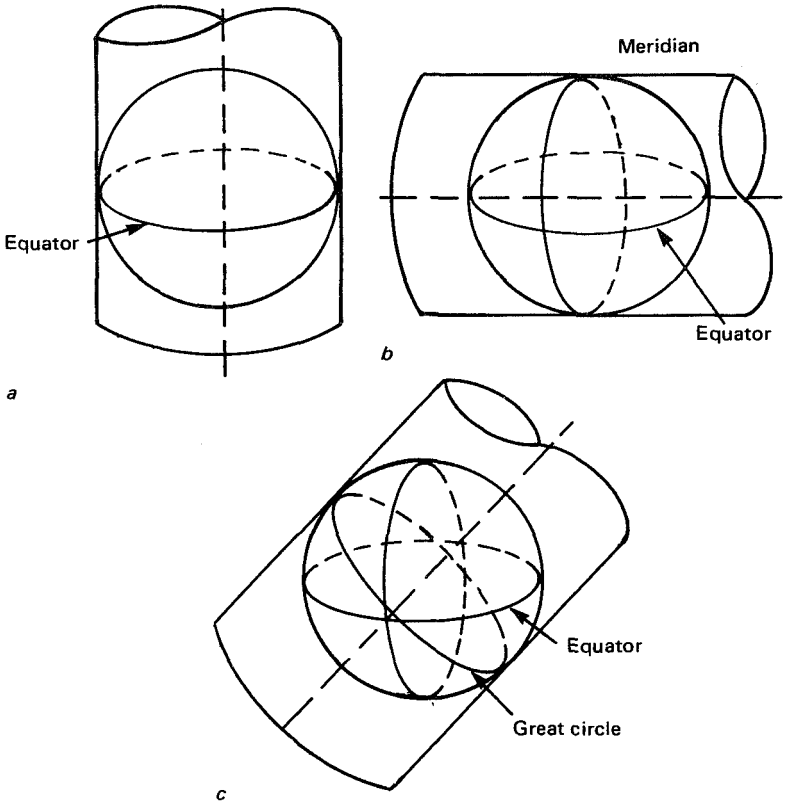


FIGURE 6.3—*Positions of projection surface. (a) Normal, contact along equator. (b) Transverse, contact along a meridian. (c) Oblique, contact along a great circle.*

point through the datum surface locate points on the projection plane. Images taken by cameras onboard spacecraft and aircraft are perspective projections if the camera axis coincides with the direction of a normal to the datum surface.

If the projection plane is tangent to the datum surface, there is no distortion at the center, and all great circles passing through the point of tangency are straight lines on the projection surface. A displacement of the projection plane along the axis changes only the scale of the projection. The location of the perspective point determines the form of the projection. Placing the perspective point diametrically opposite to the point of tangency of the projection plane with the datum surface results in a stereographic projection. If the projection axis coincides with the rotation axis of the sphere or the ellipsoid, the normal or polar stereographic

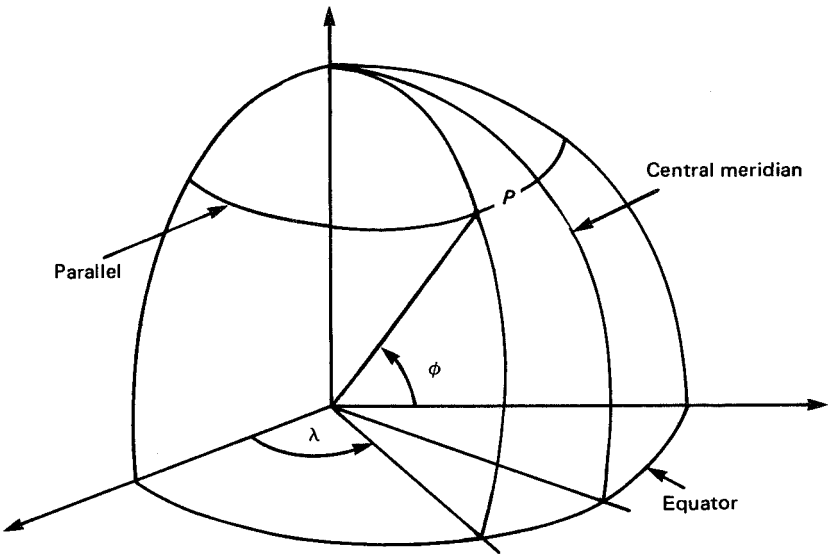


FIGURE 6.4—*Coordinate system.*

projection is obtained. In this projection the projection plane is tangent to one of the poles, with the perspective point at the other pole. The meridians are straight lines converging at the pole; the projected parallels are concentric circles about the pole. (See fig. 6.5.)

The major application of this projection is the depiction of polar areas. The transformation $\mathbf{T}_p$ for the polar stereographic projection for the sphere is given by

$$\left.\begin{aligned} x &= 2R \tan\left(\frac{\pi}{2} - \varphi\right) \sin\lambda \\ y &= 2R \tan\left(\frac{\pi}{2} - \varphi\right) \cos\lambda \end{aligned}\right\} \tag{6.2}$$

where R is the radius of the sphere.

6.3.4 Mercator Projection

The Mercator projection is a conformal cylindrical projection, with the meridians and parallels forming an orthogonal grid of straight lines. (See fig. 6.6.) The meridians are equally spaced, and the intervals between the parallels increase progressively from the equator such that the projection is conformal. By increasing the x-scale, the y-scale is matched exactly at every latitude; thus true shape is maintained. This conformality

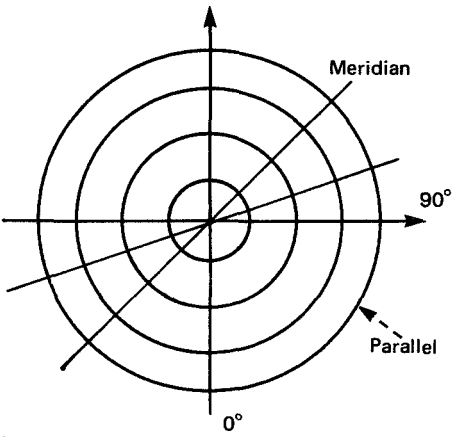


FIGURE 6.5—*Polar stereographic meridians and parallels.*

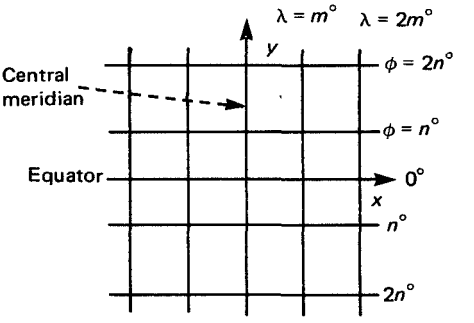


FIGURE 6.6—*Mercator grid of meridians and parallels.*

means that any straight line on the Mercator projection crosses successive meridians at a constant angle, and hence is a line of constant direction (compass course, or loxodrome). In the normal Mercator projection, distances and areas are seriously exaggerated at latitudes greater than 40°. The transformation $\mathbf{T}_p$ for the normal Mercator projection for the ellipsoid is

$$\left.\begin{aligned} x &= R\lambda \\ y &= R \ln\left[\tan\left(\frac{\pi}{4}+\frac{\phi}{2}\right)\left(\frac{1-E\sin\phi}{1+E\sin\phi}\right)^{E/2}\right] \end{aligned}\right\} \tag{6.3}$$

where R is the radius of the equatorial circle and E is the eccentricity of the ellipsoid.

The oblique Mercator projection is centered on any great circle other than the equator or a meridian. It has all the properties of the normal Mercator projection, except for the loxodrome property. It is usually defined for the sphere and not for the ellipsoid. Therefore, it is most useful for mapping small areas that are not oriented in the north–south direction, such as satellite and aircraft images. The angle between flight path and a meridian determines the direction of the axis of symmetry. The transformation $\mathbf{T}_p$ for the oblique Mercator projection for the sphere is

$$\left.\begin{aligned} x &= R \tan^{-1} \frac{\cos\varphi \sin(\lambda-\lambda_p)}{\sin\varphi \cos\varphi_P - \cos\varphi \sin\varphi_p \cos(\lambda-\lambda_p)} \\ y &= \frac{R}{2} \ln \frac{1+\sin\varphi \sin\varphi_p + \cos\varphi \cos\varphi_P \cos(\lambda-\lambda_p)}{1-\sin\varphi \sin\varphi_p + \cos\varphi \cos\varphi_p \cos(\lambda-\lambda_p)} \end{aligned}\right\} \quad (6.4)$$

where λ_p and φ_P are the longitude and latitude of the oblique pole, respectively.

The transverse Mercator projection uses a meridian rather than the equator as line of contact or true scale. All conformal properties of the normal Mercator projection except the loxodrome property are retained in the transverse Mercator projection. This projection is very useful for a 15° to 20° band centered on its central meridian. The transformation $\mathbf{T}_p$ is obtained from equation (6.4) for $\varphi_p = 0$:

$$\left.\begin{aligned} x &= R \tan^{-1} [\cos\varphi \sin(\lambda-\lambda_p)] \\ y &= \frac{R}{2} \ln \frac{1+\cos\varphi \cos(\lambda-\lambda_p)}{1-\cos\varphi \cos(\lambda-\lambda_p)} \end{aligned}\right\} \quad (6.5)$$

6.3.5 Lambert Projection

The Lambert conical projection is a conformal projection. The apex of the cone lies on the rotational axis of the ellipsoid or sphere. Meridians are straight lines converging at the apex, which is also the center of all projected circular parallels. In the secant Lambert projection the cone intersects the datum surface at two standard parallels φ_1 and φ_2. The condition that there be no distortion at the two standard parallels determines the latitude φ_0 of the central parallel circle. The intersection of the central meridian and the central parallel is the origin of the Cartesian coordinate system (x, y) in the projection plane with the y axis along the control meridian (fig. 6.7).

The transformation $\mathbf{T}_p$ for the Lambert normal conical projection for the sphere with two standard parallels φ_1 and φ_2 is

$$\left.\begin{aligned} x &= \rho \sin\theta \\ y &= \rho_0 - \rho \cos\theta \end{aligned}\right\} \quad (6.6)$$

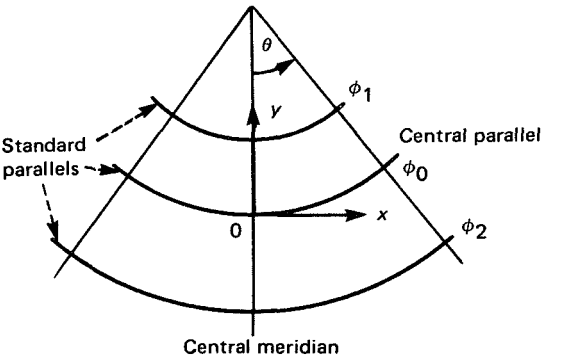


FIGURE 6.7—*Conical projection with two standard parallels.*

where

$$\rho = R \frac{\cos \varphi_1}{\sin \varphi_0} \left[\frac{\tan\left(\frac{\pi}{4} - \frac{\varphi}{2}\right)}{\tan\left(\frac{\pi}{4} - \frac{\varphi_1}{2}\right)} \right]^{\sin \varphi_0}$$

$$\theta = \lambda \sin \varphi_0$$

$$\sin \varphi_0 = \frac{\ln \cos \varphi_1 - \ln \cos \varphi_2}{\ln \tan\left(\frac{\pi}{4} - \frac{\varphi_1}{2}\right) - \ln \tan\left(\frac{\pi}{4} - \frac{\varphi_2}{2}\right)}$$

The scale distortion is dependent only on the latitude φ and not the longitude λ. Therefore, the scale distortion of a parallel circle is constant, making the Lambert conical projection suitable for areas extended in an east–west direction.

6.3.6 Universal Transverse Mercator Projection

The universal transverse Mercator (UTM) projection is actually not a projection, but a grid system based on the transverse Mercator projection. Central meridians are constructed every 6° of longitude, extending from 80° north to 80° south. Thus 60 zones, extending 3° to either side of each central meridian, are defined, and each zone is overlaid by a rectangular grid. A scale distortion or grid scale constant of 0.9996 is applied along the central meridian of each zone to reduce scale distortion of the projection. The effect is that the transverse cylinder is secant to the datum surface instead of tangent.

Whereas in the ordinary transverse Mercator projection there is no scale distortion along the central meridian, and small circles parallel to it are represented by vertical lines with increasing scale distortion away from

the central meridian, in the UTM there are two standard meridians, and the scale distortions are more evenly spread over the zone. Surface coordinates are measured in meters from the central meridian and from the equator. The central meridian is assigned a bias of 500,000 m to maintain positive values over the zone. Distances perpendicular to the central meridian are added or subtracted from this value and are called easting values. For the Southern Hemisphere a bias of 10 million m is assigned to the equator, and the northing coordinate is found by subtracting the distance to the equator from the bias value. In the Northern Hemisphere northing is simply the distance north of the equator in meters. The northing and easting coordinates, together with the zone number, define locations on the Earth within the UTM system. Polar areas are excluded from the UTM system.

6.4 Map Projection of Images

Remotely sensed images are mappings of a curved surface onto a plane and therefore contain the distortions of a map projection. Furthermore, the images are subject to the geometric distortions discussed in section 2.4.1. The transformation of an image to a map projection involves basically two steps. First, the relationship between point locations (l, s) in the distorted input image and geodetic coordinates (latitude φ and longitude λ) must be established:

$$(\lambda, \varphi) = \mathbf{T}_c(l, s) \tag{6.7}$$

Second, with the equations of the desired map projection, the x, y coordinates of the points, the projection plane, must be computed:

$$(x, y) = \mathbf{T}_p(\lambda, \varphi) \tag{6.8}$$

Finally, the projection plane coordinates must be scaled to produce an output picture with M rows and N columns,

$$(L, S) = \mathbf{T}_s(x, y) \tag{6.9}$$

where $\mathbf{T}_c$, $\mathbf{T}_p$, and $\mathbf{T}_s$ are vector functions. The composite mapping from the distorted input to the projected output image is given by

$$(L, S) = \mathbf{T}_s\mathbf{T}_p\mathbf{T}_c(l, s) = \mathbf{T}(l, s) \tag{6.10}$$

The coordinate systems used are shown in figure 6.8. The origin of the input space is the upper left corner of the input image (l, s). The origin of the projection plane or tangent space (x, y) is the image nadir point. The origin of the output space is the upper left corner of the output image.

In practice, calculation of the exact location of each image point would require a prohibitive amount of computer time. Depending on the nature of the geometric distortions and the chosen map projection, points in the projected image may be sparsely and not equally spaced. To obtain a

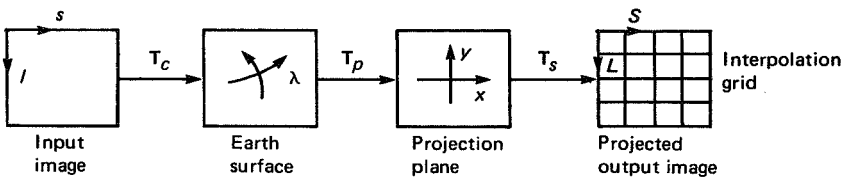


FIGURE 6.8—*Map projection coordinate systems.*

continuous picture with equally spaced elements in a reasonable time, the inverse approach is taken. A set of tie points defining a rectangular or quadrilateral interpolation grid in the output image is selected. The exact mapping transformation is only computed for the grid points. The locations of points within each quadrilateral are determined by bilinear interpolation between the vertex coordinates. Values for fractional pixel locations in the input picture are determined by one of the resampling interpolation schemes. (See sec. 3.3.2.)

The latitude and longitude for each grid point $(L, S)_G$ are determined with the inverses of $\mathbf{T}_s$ and the map projection $\mathbf{T}_p$

$$(\lambda, \varphi)_G = \mathbf{T}_p^{-1}\,\mathbf{T}_s^{-1}(L, S)_G \tag{6.11}$$

The relationship between $(\lambda, \varphi)_G$ and the input image grid coordinates $(l, s)_G$ is given by

$$(l, s)_G = \mathbf{T}_c^{-1}(\lambda, \varphi)_G \tag{6.12}$$

where $\mathbf{T}_c$ describes the viewing geometry of the imaging system. The form of $\mathbf{T}_c$ depends on the optical characteristics of the sensor, the shape and size of the datum surface, and the position and attitude of the sensor [3]. In scanning imaging systems each pixel is obtained at a different time [4], and scanner images may be considered to approximate one-dimensional perspective projections of the object scene. Often the attitude of the sensor is either not available or only inaccurately given, and the transformation $\mathbf{T}_c$ can not be calculated from a priori information. Therefore, another approach, based on the displacement of ground control points (GCPs), is used to determine $\mathbf{T}_c$. GCPs are recognizable geographic features or landmarks whose actual geodetic positions can be measured in existing maps. The coordinates $(l, s)_R$ of GCPs in the input image may be determined from shade prints or by a cross-correlation technique if a library of GCP templates is available. (See sec. 5.3 for registration techniques.)

With the assumption that the transformation from input image to projection plane $\mathbf{T}_p\mathbf{T}_c$ can be represented with sufficient accuracy by a bivariate polynomial of degree m:

$$x=\sum_{j=0}^{m}\sum_{k=0}^{m-j} a_{jk}\, l^j s^k$$
$$y=\sum_{j=0}^{m}\sum_{k=0}^{m-j} b_{jk}\, l^j s^k \tag{6.13}$$

The inverse $(\mathbf{T}_p\mathbf{T}_c)^{-1}$ is also given by a bivariate polynomial of the same degree.

Let the coordinates of the given ground control points in the geodetic coordinate system and the projection plane be $(\lambda, \varphi)_R$ and $(x, y)_R$, respectively. The following approach to produce map projections of remotely sensed images is used [5]:

1. Choose an appropriate map projection $\mathbf{T}_p$ and determine
$$(x, y)_R=\mathbf{T}_p(\lambda, \varphi)_R \tag{6.14}$$
2. Coordinates $(x, y)_R$ and $(l, s)_R$ are related by
$$(x, y)_R=\mathbf{T}_p\mathbf{T}_c(l, s)_R$$
Determine the coefficients of the polynomials representing $\mathbf{T}_p\mathbf{T}_c$ and $(\mathbf{T}_p\mathbf{T}_c)^{-1}$ by least squares by using the ground control point coordinates.
3. Determine the extent of the projection plane, circumscribe a rectangle, and scale to output image coordinates (L, S) with $\mathbf{T}_s$.
4. Divide the output image into an interpolation grid $(L, S)_G$ and determine grid point locations in input image
$$(l, s)_G=(\mathbf{T}_p\mathbf{T}_c)^{-1}\,\mathbf{T}_s^{-1}(L, S)_G \tag{6.15}$$
5. Perform a geometric transformation of the input image by using interpolation grid coordinates $(L, S)_G$, $(l, s)_G$, and a selected resampling algorithm. The actual geometric transformation is described in section 3.3.

6.5 Creating Digital Image Overlays

Overlays of remotely sensed images with maps and other images are required for change detection, map generation and updating, and modeling with multisensor and multitemporal data. Depending on the requirements, overlays may be generated with respect to a standard map projection or to a reference frame.

The generation of image-map overlays is illustrated in figures 6.9 and 6.10. Figure 6.9*a* shows an unprocessed Landsat MSS 5 image of the Baton Rouge, La., area. A UTM map of the same area is shown in figure 6.9*b*. A UTM and a normal Mercator projection of the Landsat MSS image are shown in figures 6.10*a* and 6.10*b*, respectively. Ninety-five ground control points were available to determine the coefficients of a fourth-order polynomial representing the transformation $(\mathbf{T}_p\mathbf{T}_c)^{-1}$. The transformation $\mathbf{T}_s$ was determined such that the pixel size in the projected image is 100 by 100 m. The geometric transformation to the map projection was performed by using an interpolation grid of 10 by 10 rectangular areas.

An application of image overlays for change detection is shown in figure 6.11. Two Landsat MSS scenes centered near Cairo, Ill. are used to observe the effects of spring flooding in the Mississippi Valley [6]. The extracted subimage shows areas affected by high water.

6.6 Creating Digital Image Mosaics

Mosaicking permits the analysis of remotely sensed images across frame boundaries. Depending on the requirements, the images may be mosaicked with respect to a standard map projection or to a reference frame. In the latter case a distinction can be made between mosaicking of image frames from the same or from different orbits. Adjacent frames from the same orbit exhibit fairly consistent geometric distortions, and, therefore, geometrically correct mosaics can usually be obtained by simple translation of one image with respect to the other. The attitude changes of the sensor at different orbits, however, cause geometric distortions that make it impossible to create geometrically correct mosaics of frames from different orbits without geometric rectification. This step requires that the frames to be mosaicked share a sufficiently large region of overlap.

In addition to geometric distortions, there are intensity differences that cause artificial edges at the seam between the frames. These intensity differences are due to changes in atmospheric transmittance and in illumination caused by different Sun angles. Seasonal changes of surface reflectance, precipitation, and changes caused by human activities also contribute to artificial edges in mosaics and thus interfere with image analysis. Figure 6.12, a mosaic of two unprocessed Landsat MSS 5 frames from different orbits, shows pronounced artificial edges along the vertical and horizontal seams. The geometric distortions in the overlap region of the two frames are shown in figure 6.13.

A first-order correction consists of adjusting the average gray level of each image to the same value. This preprocessing operation is in general not sufficient to eliminate the artificial edges. An improvement may be achieved by selecting subareas in the overlap region and determining a

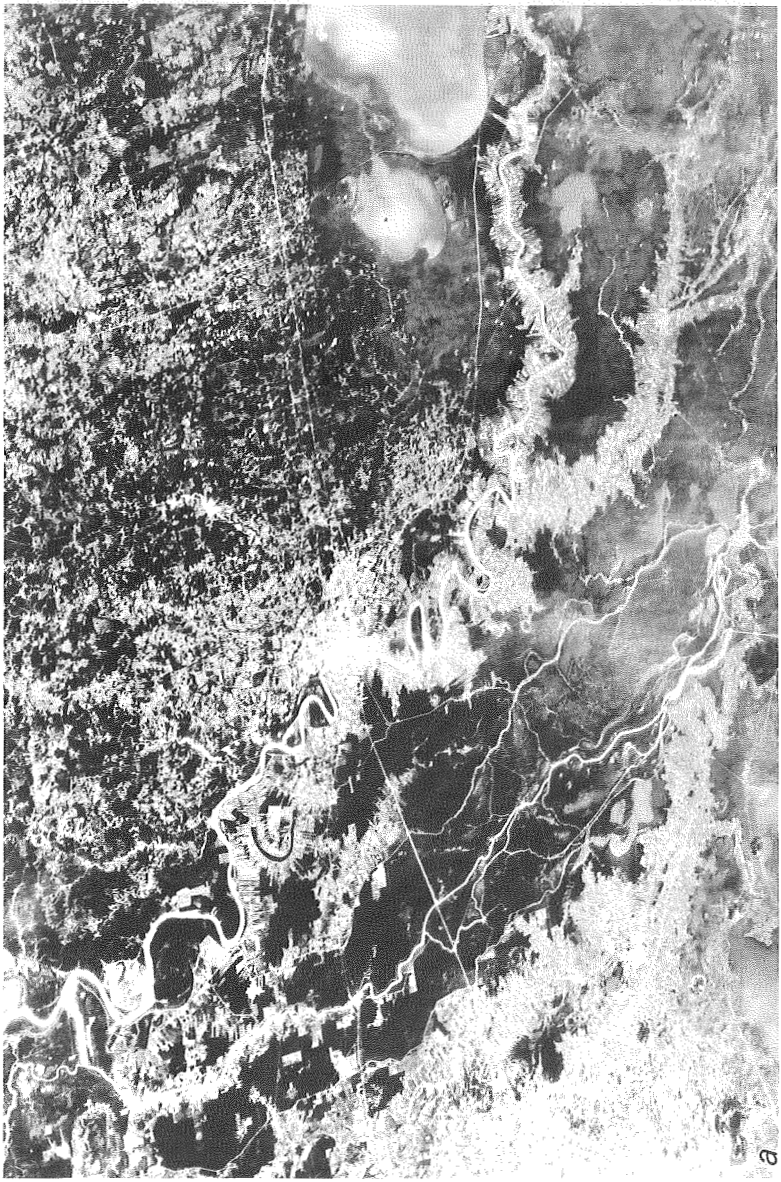


FIGURE 6.9a—*Creation of image-map overlay. Unprocessed Landsat MSS 5 image.*

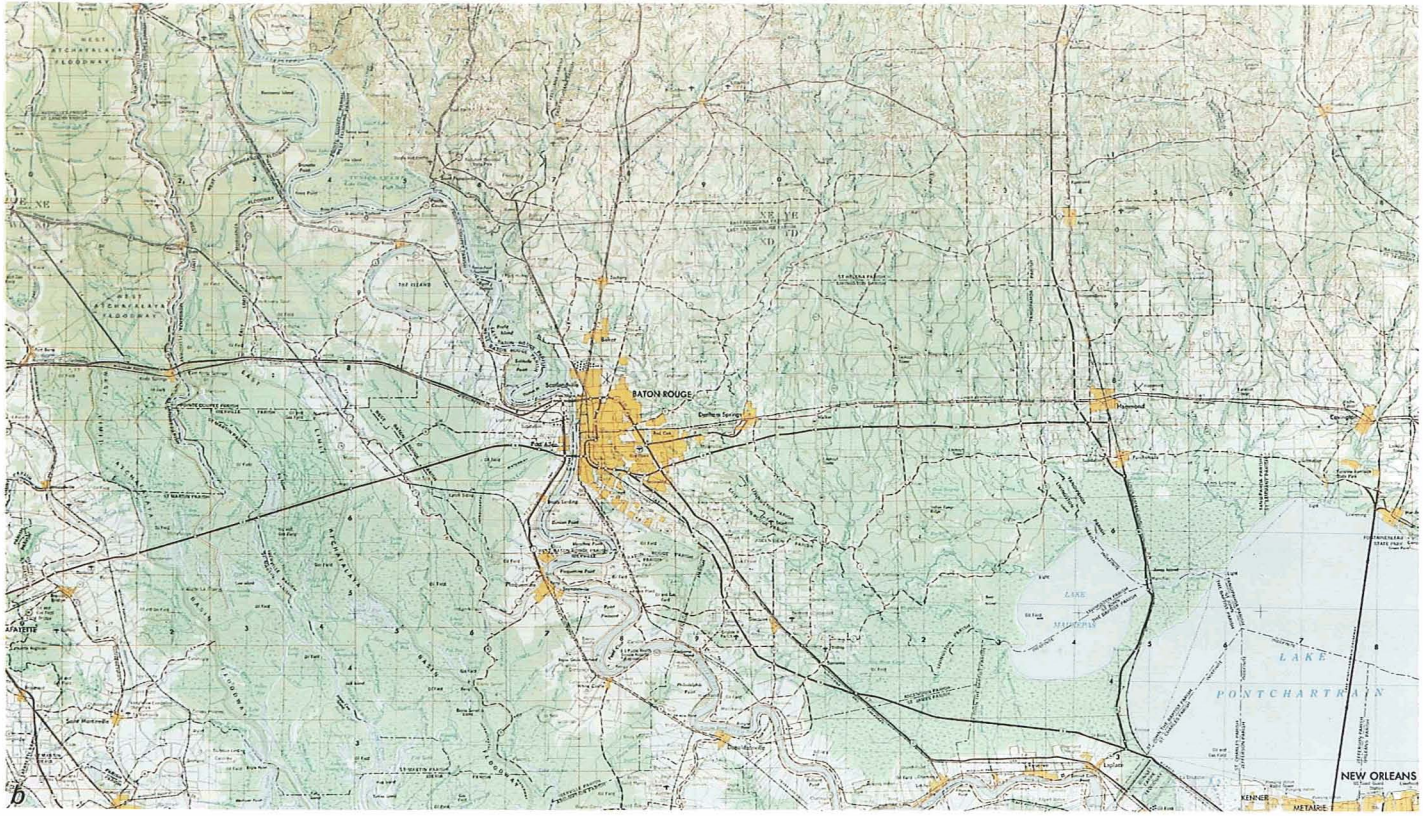


FIGURE 6.9*b*—*UTM map of area in image in part a.*

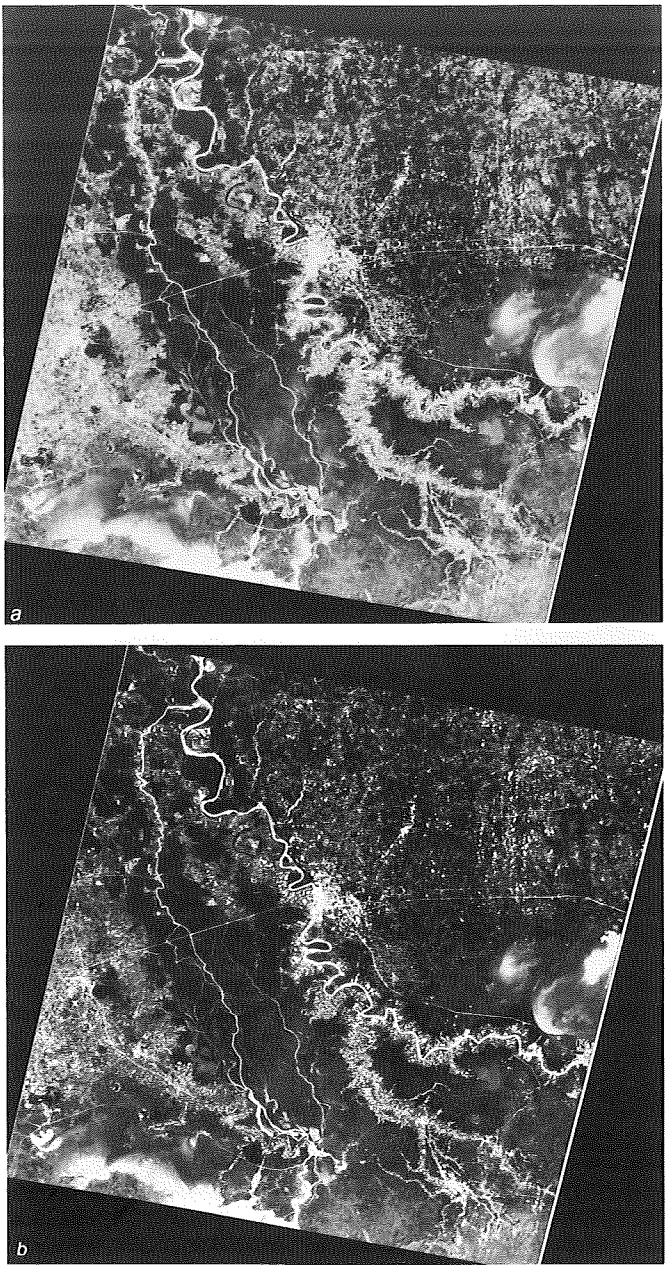


FIGURE 6.10—*Map projections of image in figure 6.9a. (a) UTM projection. (b) Normal Mercator projection.*

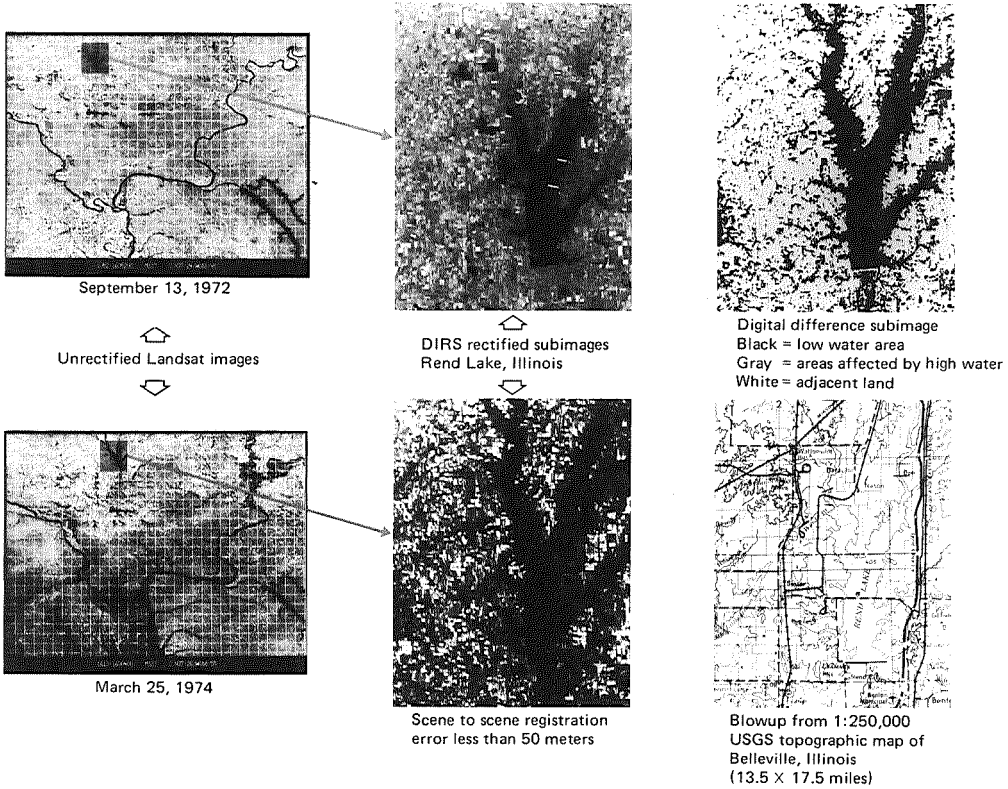


FIGURE 6.11—*Image overlay for change detection. Seasonal surface water variation determined from digitally rectified Landsat data. (Image courtesy of P. VanWie, NASA Goddard Space Flight Center).*

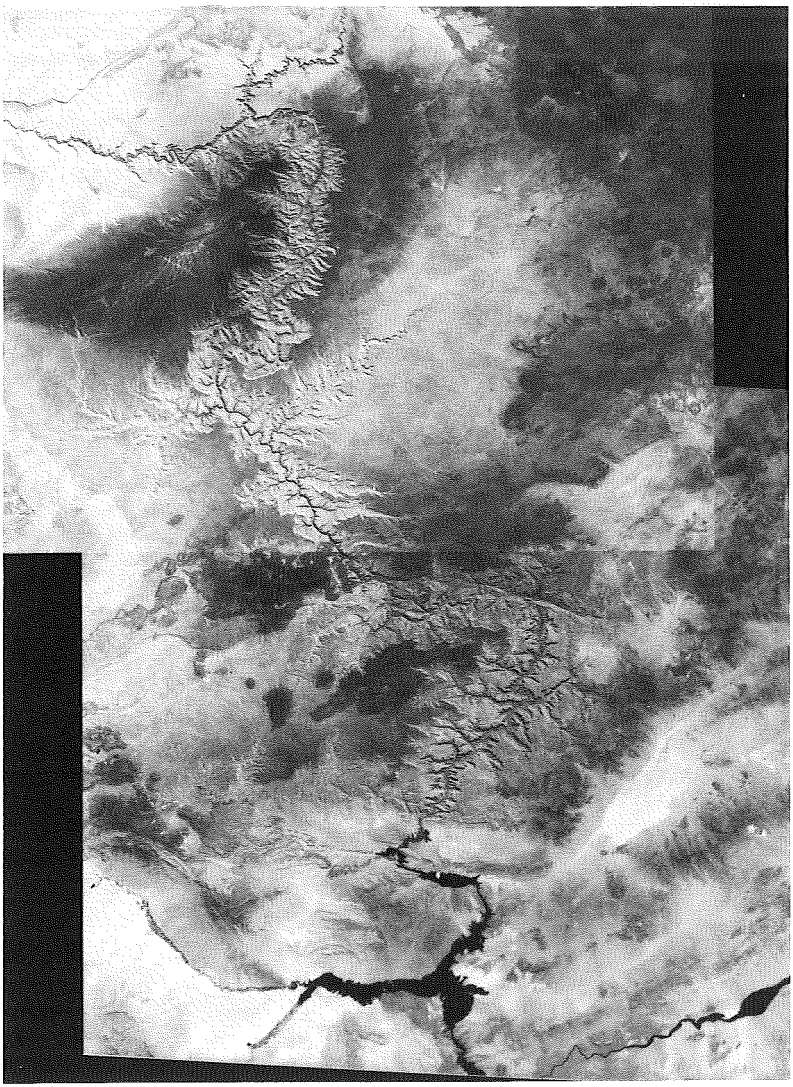


FIGURE 6.12—*Mosaic of two unprocessed Landsat MSS 5 frames of Grand Canyon.*

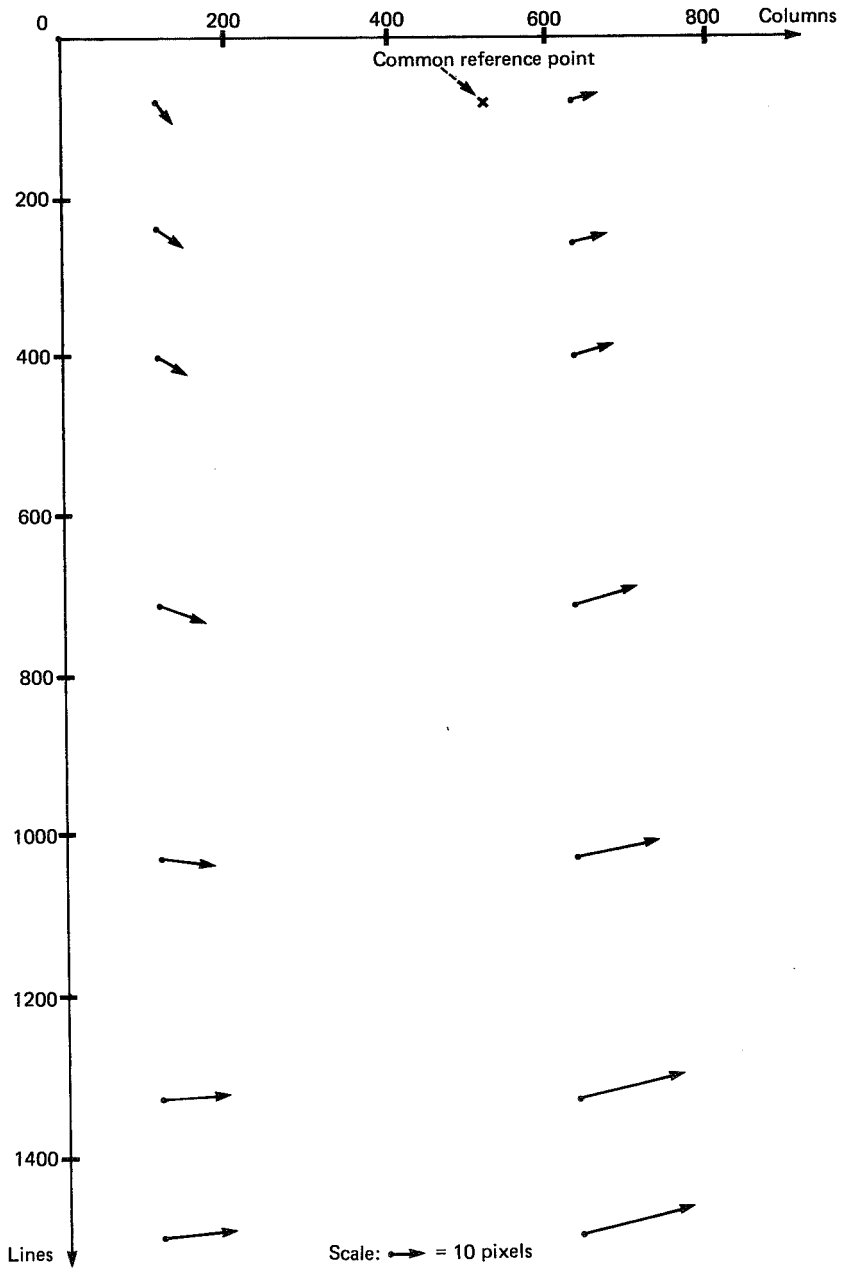


FIGURE 6.13—*Relative geometric distortions in overlap region of the mosaic in figure 6.12.*

linear two-dimensional gray-scale transformation that matches the average gray levels in the subareas. Unless there are severe brightness changes, e.g., clouds and snow fields, this technique eliminates most of the artificial edges along the vertical and horizontal seams. A further improvement is possible by finding the seam point on each line that causes the minimal artificial edge [7]. Let g_1 and g_2 be the two images to be mosaicked, and let K be the width of the overlap region. For the definition of a vertical seam, the best seam point on line j is chosen at the location where the sum of gray-value differences over a neighborhood of L pixels on the same line is minimized. That is, the seam point k^* is determined such that

$$\sum_{l=0}^{L-1} |g_1(j, k+l) - g_2(j, k+l)| = \text{minimum} \qquad k=1, \ldots, K-L+1 \tag{6.16}$$

The gray-level difference at the seam point may be smoothed by interpolation. Better results may be achieved by two-dimensional seam definition and smoothing and by extension to the spectral dimension if mosaicking of multispectral images is required.

Figure 6.14 shows a digital mosaic of parts of three Landsat MSS frames created by geometric correction, linear brightness adjustment, seam definition with equation (6.16), and seam smoothing. Figure 6.15 is a false-color mosaic of the same area.

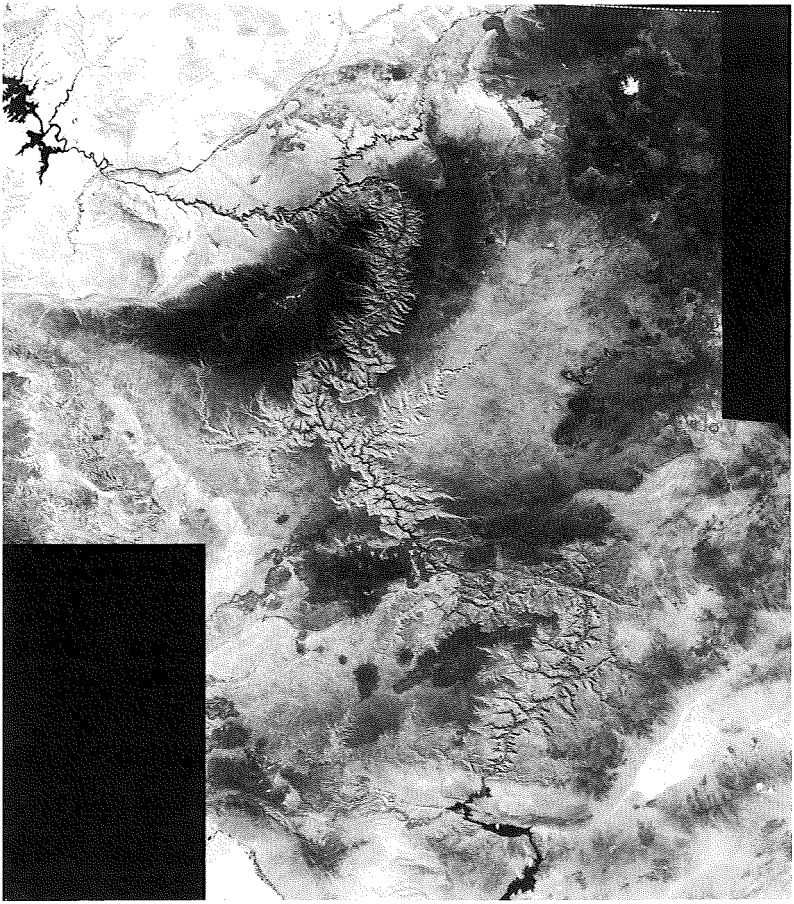


FIGURE 6.14—*Digital mosaic of parts of three Landsat MSS frames of the Grand Canyon area.*

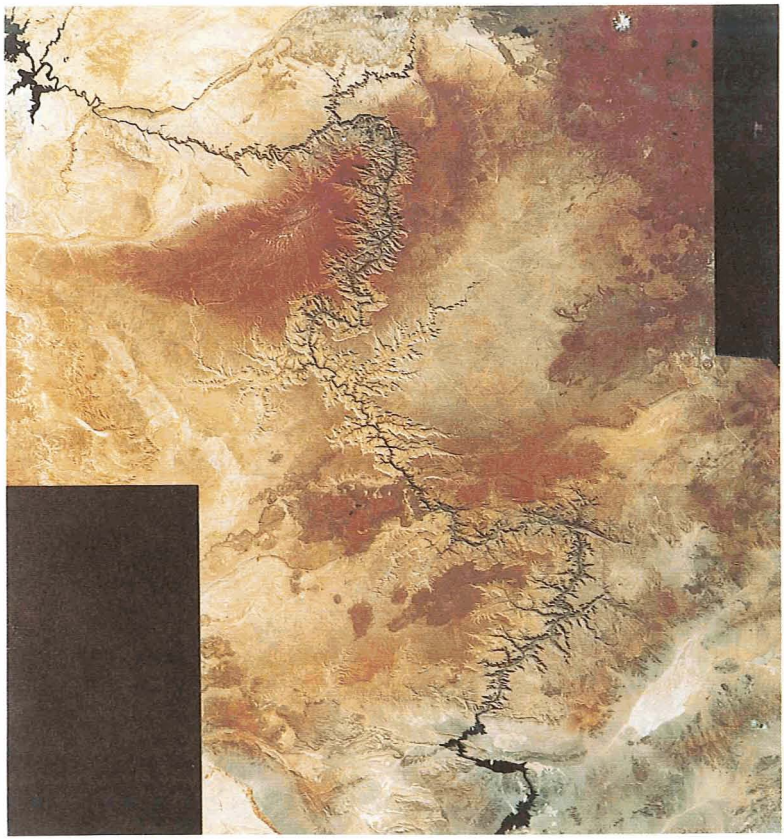


FIGURE 6.15—*False-color mosaic of Grand Canyon area.*

REFERENCES

[1] Richardus, R.; and Adler, R. K.: Map Projections. North-Holland/American Elsevier, Amsterdam, London, and New York, 1972.
[2] Gilbert, E. N.: Distortion in Maps, SIAM Rev., vol. 16, no. 1, 1974, pp. 47–62.
[3] Elliot, D. A.: Digital Cartographic Projection. Proceedings of Caltech/JPL Conference on Image Processing Technology, Data Sources and Software for Commercial and Scientific Applications, California Institute of Technology, Pasadena, Calif., Nov. 1976, pp. 5-1–5-10.
[4] Puccinelli, E. F.: Ground Location of Satellite Scanner Data, Photogr. Eng. and Remote Sensing, vol. 42, 1976, pp. 537–543.
[5] Moik, J. G.: Smips/VICAR Application Program Description, NASA TM 80255, 1979.
[6] Van Wie, P.; and Stein, M.: A Landsat Digital Image Rectification System, IEEE Trans. Geosci. Electron., vol. GE-15, 1977, pp. 130–137.
[7] Milgram, D. L.: Computer Methods for Creating Photomosaics, IEEE Trans. Comput., vol. c-24, 1975, pp. 1113–1119.

7. Image Analysis

7.1 Introduction

Image analysis is concerned with the description of images in terms of the properties of objects or regions in the images and the relationships between them. Although image restoration and enhancement produce images again, the result of image analysis operations is a description of the input image. This description may be a list describing the properties of objects such as location, size, and shape; a relational structure; a vector field representing the movement of objects in a sequence of images; or a map representing regions. In the latter case the description is again pictorial, but its construction requires the location of regions and determination of their shapes.

The description always refers to specific parts in the image. Therefore, to generate the description, it is necessary to segment the image into these parts. The parts are determined by their homogeneity with respect to a given gray-level property such as constant gray value and texture, or a geometric property based on connectedness, size, and shape [1]. Thus, image analysis involves image segmentation and description of the segmented image in terms of properties and relationships.

7.2 Image Segmentation

Image segmentation is that part of image analysis that deals with the spatial definition of objects or regions in an image. Objects have two basic characteristics: (1) They exhibit some internal uniformity with respect to an image property, and (2) they contrast with their surroundings. Because of noise, the nature of these characteristics is not deterministic. One property is gray level, because many objects are characterized by constant reflectance or emissivity on their surface. Thus, regions of approximately constant gray level indicate objects. Another property is texture, and regions of approximately uniform texture may represent objects.

A region R_i is a set of points surrounded by a closed curve of finite length. Regions have the property of being simply connected. A segmenta-

tion of the image domain R is a finite set of regions $(R_1, R_2, \ldots, R_I)$ such that

$$\left.\begin{aligned} R &= \bigcup_{i=1}^{I} R_i \\ R_j \bigcap R_i &= \varnothing \qquad \text{for } j \neq i \end{aligned}\right\} \tag{7.1}$$

where $\varnothing$ is the empty set and $\cup$ and $\cap$ represent the set operations union and intersection, respectively. (See fig. 7.1.)

Image segmentation can be obtained on the basis of both regional and border properties. Given a regional property such as intensity, color distribution, or texture, picture elements that are similar with respect to this property may be combined into regions. Alternatively, the borders between regions may be located by detecting discontinuities in image properties.

An image property is a function that maps images into numbers. The value of the property for a given image g is the number obtained by the operation. Examples of image properties are: (1) The gray level $g(j_0, k_0)$ of g at a given point (j_0, k_0); (2) the average gray level of a neighborhood of (j_0, k_0); (3) the coefficients of an orthogonal transformation (e.g., Fourier and Karhunen–Loéve); and (4) geometrical properties such as connectedness, area, and convexity. Property 2 is a local image property.

7.2.1 Thresholding

Gray-level thresholding is an effective and simple segmentation technique when the objects have a characteristic range of gray levels. If g is a single-component image with gray-level range $[z_1, z_k]$ that contains I

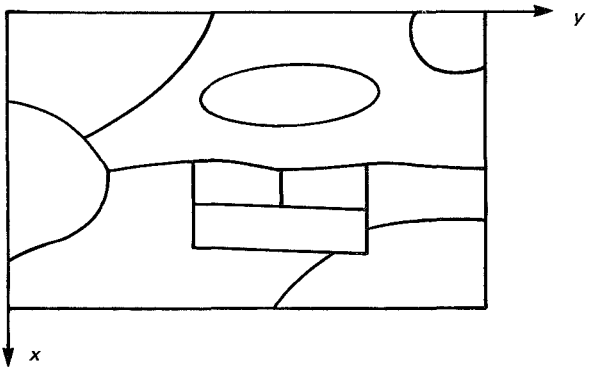


FIGURE 7.1—*Segmentation of image into regions.*

regions with the nonoverlapping gray-level ranges $Z_i \subset [z_1, z_k]$, $i=1, \ldots, I$, then a threshold image g_t is defined by

$$g_t(j, k) = \begin{cases} i & \text{if } g(j, k) \in Z_i \\ 0 & \text{otherwise} \end{cases} \tag{7.2}$$

The histogram of gray levels is examined, and if it is strongly multimodal, it can be assumed that the image contains approximately uniform areas that constitute regions [1, 2]. Thus, the image can be segmented by bringing it to a threshold at the lowest gray levels between histogram peaks. Threshold selection is facilitated in an interactive system by repeatedly displaying the histogram and evaluating the result of such selection. Thresholding can also be used to segment an image into regions of uniform texture. A given image g is transformed into an image g_t by computing a local texture property at every image point (see sec. 7.4) and bringing to a threshold.

7.2.2 Edge Detection

Edge detection is an image segmentation method based on the discontinuity of gray levels or texture at the boundary between different objects. Such a discontinuity is called an edge. An edge separates two regions of relatively uniform but different gray level or texture. Another type of gray-level discontinuity is the line; it differs from the regions on both sides. Edge detection involves context-free algorithms that make no assumptions about edge continuity. A common approach to edge detection in monochrome images is edge enhancement (see sec. 4.3) followed by a thresholding operation to determine the locations of significant edges.

Classical edge detectors are derivative operators, which give high values at points where the gray level of the image g changes rapidly [1, 3–5]. For digital images, differences rather than derivatives are used. The first-order differences in the x and y directions are

$$\Delta_x g(j, k) = g(j, k) - g(j-1, k) \tag{7.3}$$

$$\Delta_y g(j, k) = g(j, k) - g(j, k-1) \tag{7.4}$$

First differences in other directions θ can be defined as linear combinations of the x and y differences as

$$\Delta_\theta g(j, k) = \Delta_x g(j, k) \cos\theta + \Delta_y g(j, k) \sin\theta \tag{7.5}$$

With the magnitude of the maximum directional difference, the gradient, an edge-enhanced image g_e is obtained by

$$g_e(j, k) = \sqrt{[\Delta_x g(j, k)]^2 + [\Delta_y g(j, k)]^2} \tag{7.6}$$

The magnitude of the gradient, as given in equation (7.6), detects edges in all orientations with equal sensitivity. Equation (7.6) is often approximated by

$$g_e \approx |\Delta_x g(j, k)| + |\Delta_y g(j, k)| \tag{7.7}$$

or by

$$g_e \approx \max [|\Delta_x g(j, k)|, |\Delta_y g(j, k)|] \tag{7.8}$$

These approximations are no longer equally sensitive to edges in all orientations. Various other edge-enhancement operators can be defined [3, 6]. Let

$$d_1 = |g(j, k+1) - g(j, k-1)| \tag{7.9}$$

$$d_2 = |g(j+1, k) - g(j-1, k)| \tag{7.10}$$

$$d_3 = |g(j+1, k+1) - g(j-1, k-1)| \tag{7.11}$$

$$d_4 = |g(j-1, k+1) - g(j+1, k-1)| \tag{7.12}$$

$$d_5 = |g(j-1, k+1) + 2g(j, k+1) + g(j+1, k+1) - g(j-1, k-1) - 2g(j, k-1) - g(j+1, k-1)| \tag{7.13}$$

$$d_6 = |g(j+1, k-1) + 2g(j+1, k) + g(j+1, k+1) - g(j-1, k-1) - 2g(j-1, k) - g(j-1, k+1)| \tag{7.14}$$

Then, edge-enhanced images are obtained by

$$g_e(j, k) = \max (d_1, d_2) \tag{7.15}$$

$$g_e(j, k) = d_1 + d_2 \tag{7.16}$$

$$g_e(j, k) = d_5 + d_6 \tag{7.17}$$

$$g_e(j, k) = d_1 + d_2 + d_3 + d_4 \tag{7.18}$$

Edges may also be enhanced by convolution of the image g with proper masks. (See sec. 4.3.)

Threshold selection is a key problem in edge detection in noisy images. Too high a threshold does not permit detection of subtle, low-amplitude edges. Conversely, setting the threshold too low causes noise to be detected as edges. Nack [6] proposed an adaptive threshold selection technique using a desired edge density D as key parameter. The normalized histogram of the enhanced image g_e is formed

$$n(z) = \frac{H_e(z)}{MN} \qquad z = 0, 1, \ldots, K-1 \tag{7.19}$$

The function $H_e(z)$ is the frequency of occurrence of gray level z in the edge-enhanced image g_e, K is the number of quantization levels, and M

and N are the image dimensions. The threshold T that determines whether a picture element in the edge-enhanced image is an edge point is calculated as

$$T=K-1-z \tag{7.20}$$

where z is determined such that the actual edge density

$$D_z=\sum_{i=0}^{z} n(K-1-i) \tag{7.21}$$

matches the desired edge density D. Varying the edge density D thickens or thins edges. The edge image $e(j, k)$, indicating the position of edges in the image g, is obtained by

$$e(j, k)=\begin{cases} 1 & g_e(j, k)\geq T \\ 0 & g_e(j, k)<T \end{cases} \tag{7.22}$$

Figure 7.2 illustrates the effects of applying different edge-enhancement operators and varying the edge density for threshold selection. Edges are represented by white pixel values against a black background. Edges were enhanced with the operators given in equations (7.15), (7.17), and (7.18), and edge densities of 5, 10, and 15 percent, respectively, were used for threshold selection. The threshold values determined by equation (7.20) are also shown. Visual evaluation indicates that the edge density is a significant parameter for edge detection, but the choice of edge-enhancement operator has little influence for this class of images. A study of edge-detector performance was reported in [7].

For multi-images, edge detection is usually performed on the component images, and the edge images obtained may be combined into a composite edge image by a logical OR operation. An important application of edge detection is in image registration where edge images are used for binary correlation. (See sec. 5.4.) Experiments have shown that for remotely sensed Earth resources images, an edge density of $D=15$ percent is appropriate.

Edge detection is of limited value as an approach to segmentation of noisy remotely sensed images. Often the edges have gaps at places where the transitions between regions are not sufficiently abrupt. Additional edges may be detected at points that are not part of region boundaries, and the detected edges will not form a set of closed, connected object boundaries. However, the object boundaries may be constructed by connecting the extracted edge elements. Thus, boundary detection is achieved by the combination of local edge detection, followed by operations that thin and link the segments obtained into continuous boundaries [8, 9].

A boundary-finding algorithm that segments multi-images into regions of any shape by merging statistically similar subregions is given in [10].

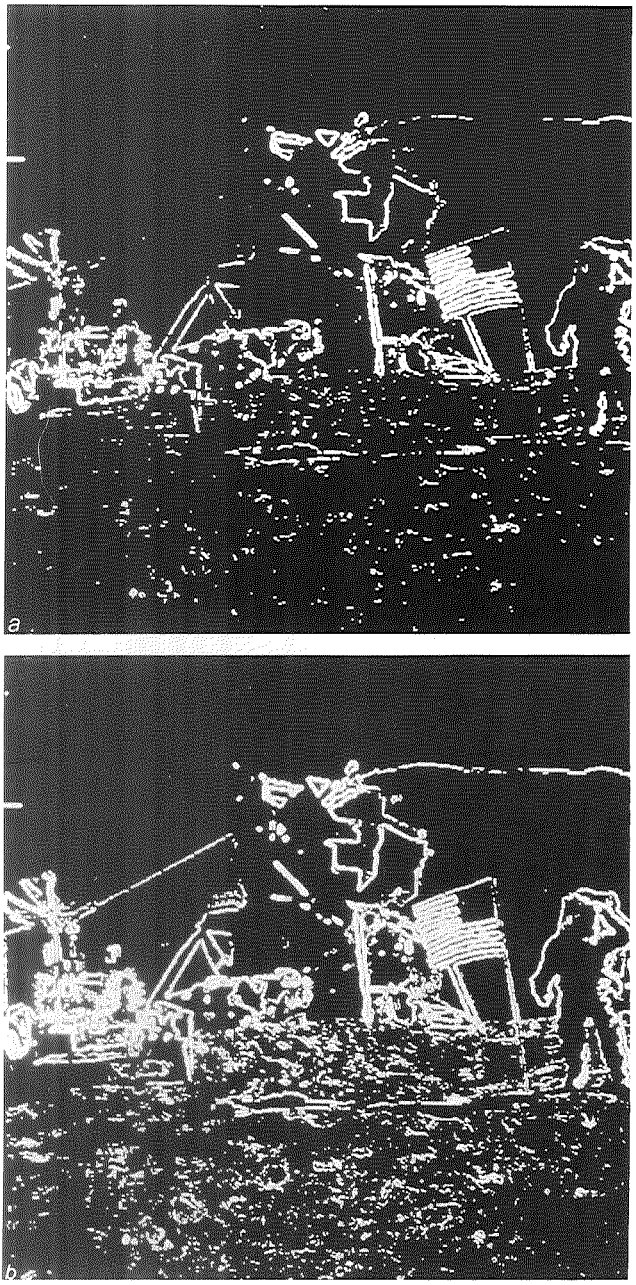


FIGURE 7.2—*Edge detection. Effects of various edge-enhancement operators T and edge densities D. Images in parts a, b, and c are enhanced with equation (7.15); parts d, e, and f, with equation (7.17); and parts g, h, and i, with equation (7.18). (a) D = 5 percent, T = 48. (b) D = 10 percent, T = 34.*

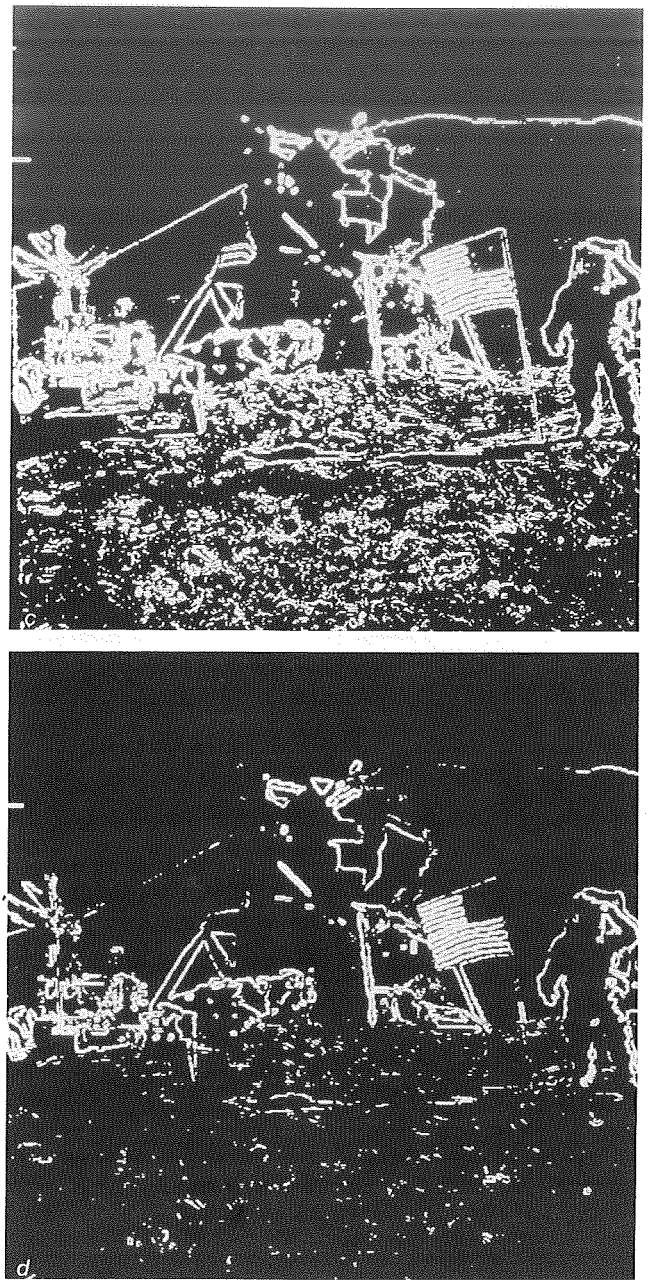


FIGURE 7.2—*Continued. (c) D = 15 percent, T = 27. (d) D = 5 percent, T = 232.*

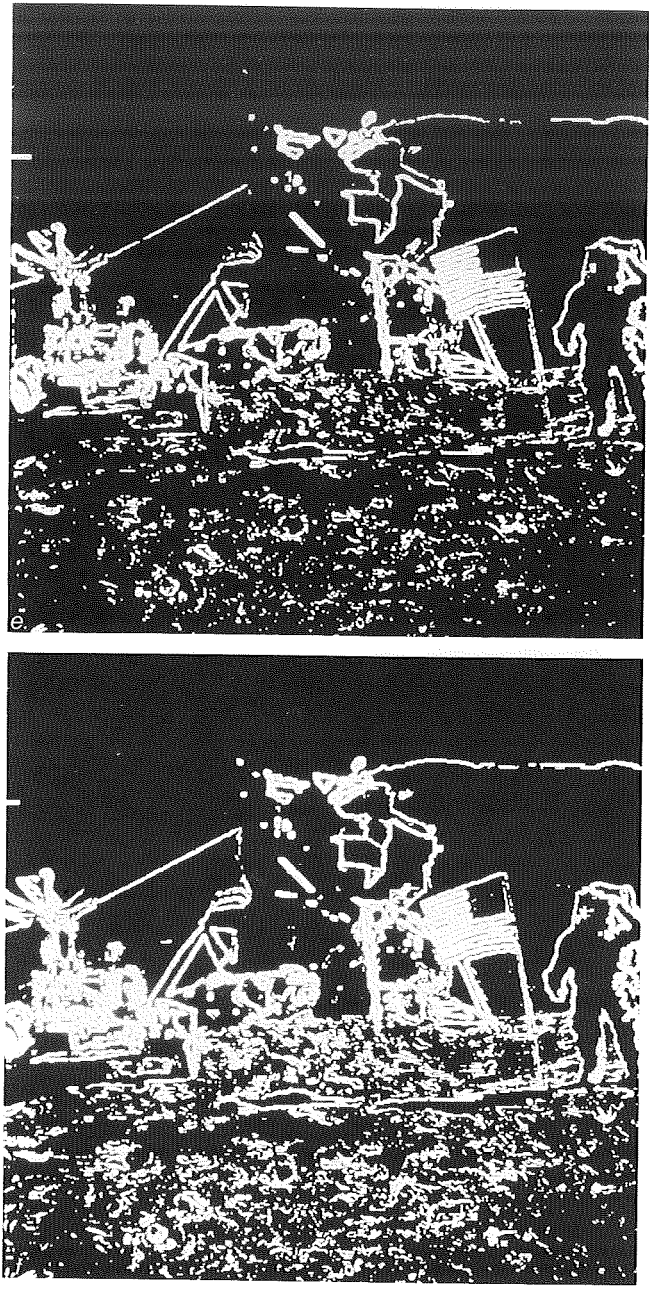


FIGURE 7.2—*Continued. (e) D = 10 percent, T = 164. (f) D = 15 percent, T = 128.*

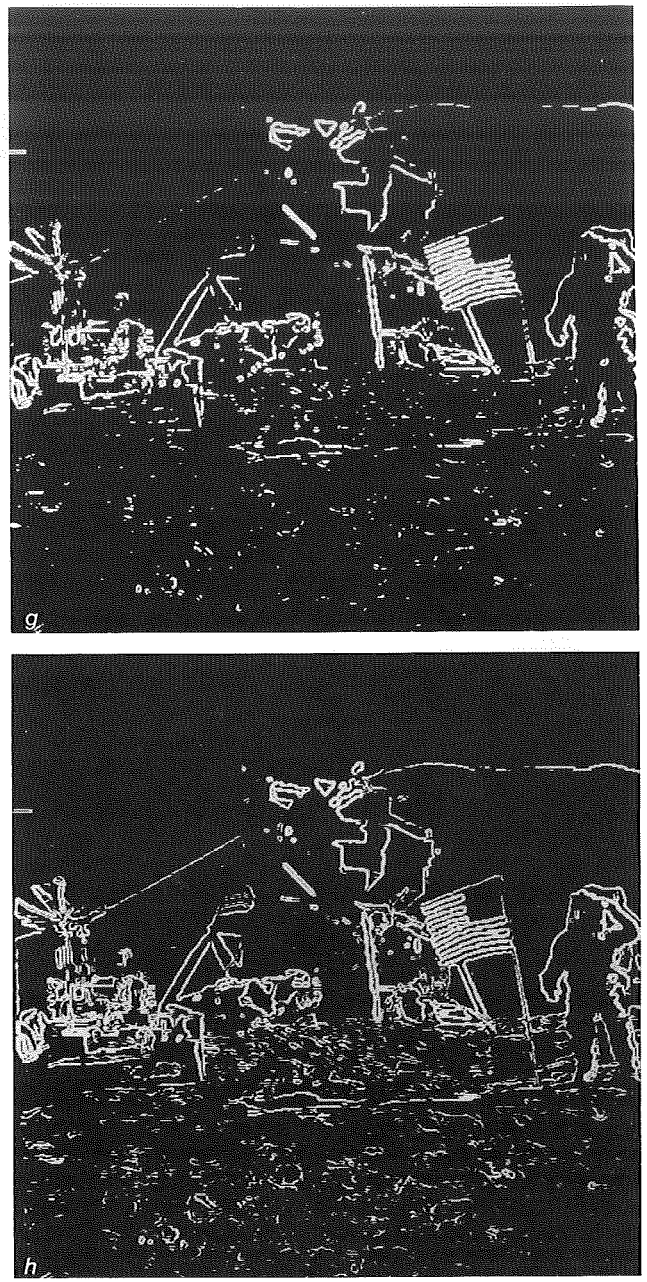


FIGURE 7.2—*Continued. (g) D = 5 percent, T = 146. (h) D = 10 percent, T = 103.*

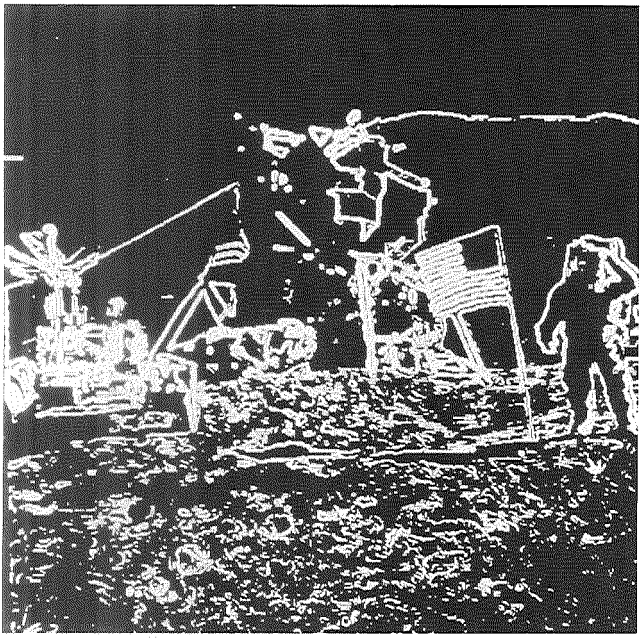


FIGURE 7.2—*Continued. (i) D = 15 percent, T = 80.*

Another technique for partitioning multispectral images into regions is described in [11].

7.2.3 Texture Analysis

By thresholding, images may be segmented into regions that are homogeneous with respect to a given image property. If objects have approximately constant reflectance over their surfaces, regions of constant gray level represent objects. More generally, regions of homogeneous texture (see sec. 2.8.3) may indicate objects [12]. Conversely, edges may not only be defined by abrupt changes in gray level but also at locations at which there is an abrupt change in texture. Thus, textural features are important for image segmentation and classification.

Current texture analysis techniques are based on Fourier expansion or statistical analysis. Hawkins [13] described texture as a nonrandom arrangement of a local elementary pattern repeated over a region that is large in comparison to the pattern's size. Texture is often qualitatively described by its coarseness.

The coarseness of a texture is related to the spatial repetition period of the local structure. A large period implies a coarse texture. Therefore, texture properties such as coarseness and directionality can be derived from the power spectrum of a texture sample. High values of the power

spectrum $|F|^2$, given in equation (2.48), near the origin indicate a coarse texture, but in a fine texture the values of $|F|^2$ are spread over much of the spatial frequency domain. The two-dimensional power spectrum also represents the directions of edges and lines in an image. A texture with a preferred direction θ will have high values of $|F|^2$ around the perpendicular direction $\theta+(\pi/2)$. Thus, texture measures can be derived from averages of the power spectrum taken over ring and wedge-shaped regions of the spatial frequency domain.

$$\phi_r = \int_0^{2\pi} |F(r, \theta)|^2 \, d\theta \tag{7.23}$$

$$\phi_\theta = \int_0^{\infty} |F(r, \theta)|^2 \, dr \tag{7.24}$$

Statistics of local image property values, such as means and variances, computed at every point of a given image g may be used as texture measures. For example, the directional differences between pairs of average gray levels were proposed as texture measures in [14].

Let $\bar{g}_m(j, k)$ be the average gray level of image g in a square region of side $m+1$ centered at (j, k). Differences of these averages, for pairs of horizontally, vertically, or diagonally adjacent local regions may be used as texture measures. The definitions of the differences are:

1. Horizontal

$$T_H(j, k) = \left| \bar{g}_m\left(j - \frac{m}{2}, k-m+1\right) - \bar{g}_m\left(j - \frac{m}{2}, k+1\right) \right| \tag{7.25}$$

2. Vertical

$$T_V(j, k) = \left| \bar{g}_m\left(j-m+1, k - \frac{2}{m}\right) - \bar{g}_m\left(j+1, k - \frac{2}{m}\right) \right| \tag{7.26}$$

3. Diagonal

$$T_{45}(j, k) = |\bar{g}_m(j-m+1, k-m+1) - \bar{g}_m(j, k)| \tag{7.27}$$

$$T_{135}(j, k) = |\bar{g}_m(j-m+1, k) - \bar{g}_m(j, k-m+1)| \tag{7.28}$$

$$m = 1, 2, \ldots$$

For a coarse texture and small displacements m of the local regions, the values in $T(j, k)$ should be small; i.e., the histogram of $T(j, k)$ should have values near zero. Conversely, for a fine texture comparable to the local region size, the elements of $T(j, k)$ should have different values so that the histogram of $T(j, k)$ is spread out.

Textural properties can also be derived from the probabilities that gray levels occur as neighbors [15]. The higher that the probability that a gray level occurs as a neighbor of the same or a similar gray level is, the finer

is the texture. Texture is a local property of a picture element. Therefore, texture measures are dependent on the size of the local observation region.

7.3 Image Description

Once an image has been segmented into regions, a description in terms of the region properties and the relationships between the regions may be obtained. Measuring properties of the regions and establishing relationships between them are often very complex processes. Rosenfeld [1] summarized the problems and emphasized that prior knowledge about the class of images under consideration should be used as a model to guide both the segmentation and the measurement of properties. Use of prior knowledge is greatly facilitated in an interactive image processing system for which the analyst combines the information available in the image with his experience from previous data.

A special case of image description is classification. Here the description is simply the name of the class to which a region or a picture element belongs. Classification has been successfully applied for the analysis of remotely sensed images and is extensively treated in chapter 8. However, classification employing statistical pattern recognition techniques uses only sets of property values of picture elements or regions to characterize an image and does not use relationships between regions. An adequate structural description of remotely sensed images has not yet been obtained because of the complexity of the images and the presence of noise. An attempt to describe the structure of a class of remotely sensed images by a web grammar with use of labeled graphs is described in [16]. Web grammars [17] provide a convenient model to represent spatial relationships. Reviews of the structural or syntactic approach to image analysis are given in [18] and [19].

Despite the lack of explicit models for remotely sensed images, complex analysis problems have been solved successfully with systems in which the knowledge about the problem domain is implicitly contained in the analysis functions and in the communication with the user. The representation of knowledge in programs is a powerful tool in the development of image analysis systems.

7.4 Image Analysis Applications

This section illustrates the successful solution of three image analysis problems, the determination of wind fields, land-use mapping, and change detection with image analysis systems that permit the use of prior knowledge about the problem at each analysis step [20, 21].

7.4.1 Wind Field Determination

This example illustrates how models implicit in the analysis functions of the Atmospheric and Oceanographic Information Processing System

(AOIPS) [20] combined with the analyst's experience are used to determine atmospheric motions from a series of satellite images. Remotely sensed images from geosynchronous satellites provide the possibility of studying the dynamics of the atmosphere [22, 23]. Atmospheric properties such as divergence, which describes horizontal atmospheric motions, can be associated with the development of severe storms and are, therefore, important for storm prediction. Divergence can be derived from a wind vector field describing the atmospheric fluid flow.

Wind vector fields may be determined by measuring cloud displacements in a series of images obtained at known time intervals. The displacement of a cloud divided by the elapsed time between the images gives the wind velocity. Accurate registration of successive images (see ch. 5) and determination of the location of the clouds with respect to the Earth are necessary to convert relative cloud displacements to wind vectors in geodetic or Earth coordinates. Therefore, the transformation between image coordinates and a reference coordinate system on the Earth must be found. Because of the lack of a perfect geosynchronous orbit of the satellite, this transformation is rather complex and time dependent. The precise attitude of the spacecraft is determined by using orbit information and fitting landmarks or ground control points. (See sec. 3.3 and 6.4.) This navigation process for spin-stabilized spacecraft was described previously [24, 25].

Thus, the first steps in estimating wind fields are to identify landmarks in a series of images covering a storm and to measure their locations in the images and a map. The AOIPS [20] provides functions to identify landmarks with a cursor, to increase the scale and the contrast of subimages surrounding landmarks, and to extract the image coordinates of landmarks. Once a sufficient set of landmarks has been defined, the navigation process is performed.

The next step is to select unique clouds in the first image of the series and to track their motion in the remaining sequence. This step requires the segmentation of the images into clouds and background. Both automatic segmentation by thresholding and manual identification of clouds with a cursor are possible. The exact location of a given cloud in subsequent images is determined by cross-correlation. (See sec. 5.3.) The problem is to select clouds for correlation that are fairly small and whose shape is approximately invariant within the sequence [26, 27]. For each tracked cloud a wind vector is determined. The cloud height is calculated by estimating the cloud optical thickness and is used to assign the level for which the cloud is tracked. The wind vectors obtained by this cloud tracking process are randomly distributed. A wind field on a uniformly spaced grid in the earth coordinate system can be determined by interpolation.

Figure 7.3*a* shows a Geostationary Operational Environmental Satellite 1/Video Infrared Spin Scan Radiometer (GOES 1/VISSR) image

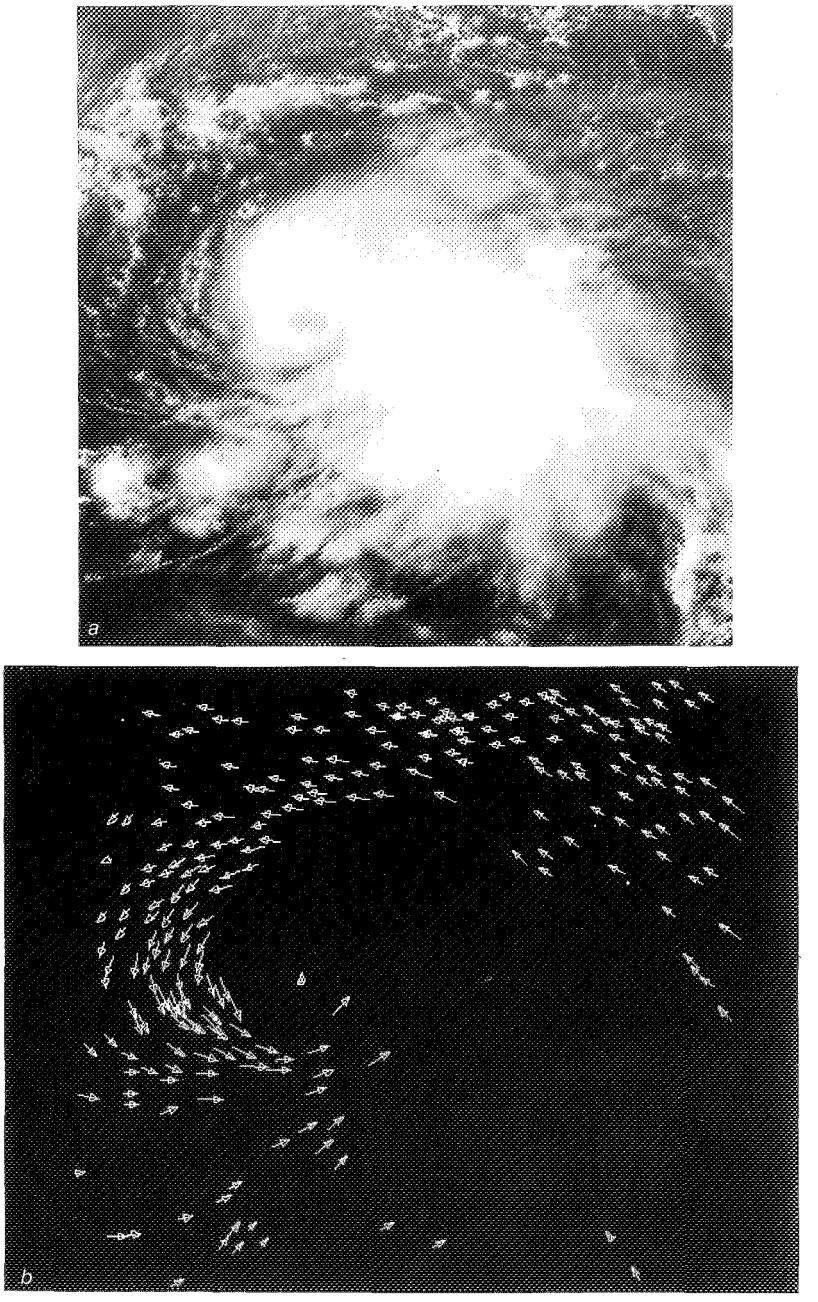


FIGURE 7.3—*Wind field determination for tropical storm Anita. (a) Visible Geostationary Operational Environmental Satellite 1/Video Infrared Spin Scan Radiometer (GOESI/VISSR) image taken on August 31, 1977, at 1600 G.m.t. (b) Lower tropospheric wind field determined from four images taken 3 min. apart.*

of tropical storm Anita obtained on August 31, 1977, over the Gulf of Mexico. A series of four images taken 3 min. apart was used to derive the wind field shown in figure 7.3*b* [28]. The length of an arrow is proportional to the wind speed. Figure 7.4 shows a visible image of the storm combined with the derived lower tropospheric wind field and with the wind field interpolated to a uniform grid. Various field parameters can be calculated from the uniform wind field. The radial and tangential wind velocity components in a polar coordinate system with the origin at the center of the storm are shown in figure 7.5. The contours represent constant values of velocity. The radial and tangential components can be used to calculate the areal horizontal mean divergence and the areal mean relative vorticity of the field, respectively. These parameters may be used as input to models for studying the dynamics of the atmosphere.

7.4.2 Land-Use Mapping

This image analysis example illustrates the description of regions that are not the result of segmentation but are defined independently and then superimposed on the remotely sensed image. Examples are urban area divisions according to population or political jurisdiction, such as census tracts or municipalities. The region boundaries are usually defined by polygonal boundaries given by a list of vertex coordinates. This spatial data structure is handled by most geographic information systems. To overlay the region boundaries on an image requires a conversion to the raster image data structure. The Image Based Information System (IBIS) [29] converts polygonal data structures to an image raster and provides functions for the registration of boundary and gray-scale images.

An important application for this combination is the integration of socioeconomic and remotely sensed data to determine land-use changes [30]. The first processing steps are to convert the polygonal data structure of the Census Bureau Urban Atlas to an image and to register the resulting boundary image to the corresponding remotely sensed image. This registration process involves the selection of a sufficient number of ground control points. (See sec. 6.4.) Here the geographic locations of ground control points are known from the Urban Atlas file and do not have to be extracted from a map as described in section 6.4. A problem, however, is the registration of tract boundaries that do not coincide with physical features in the image. Figure 7.6*a* shows the census tract boundaries of the Richmond, Va., area. These boundaries are combined with two bands of a corresponding Landsat MSS image in figure 7.6*b*. The blue lines define the original census tracts, and the yellow boundaries are registered to the image.

The next step is to segment the remotely sensed multispectral image into natural regions applying, for example, classification. Figure 7.6*c* shows the classification map obtained by a clustering technique. (See

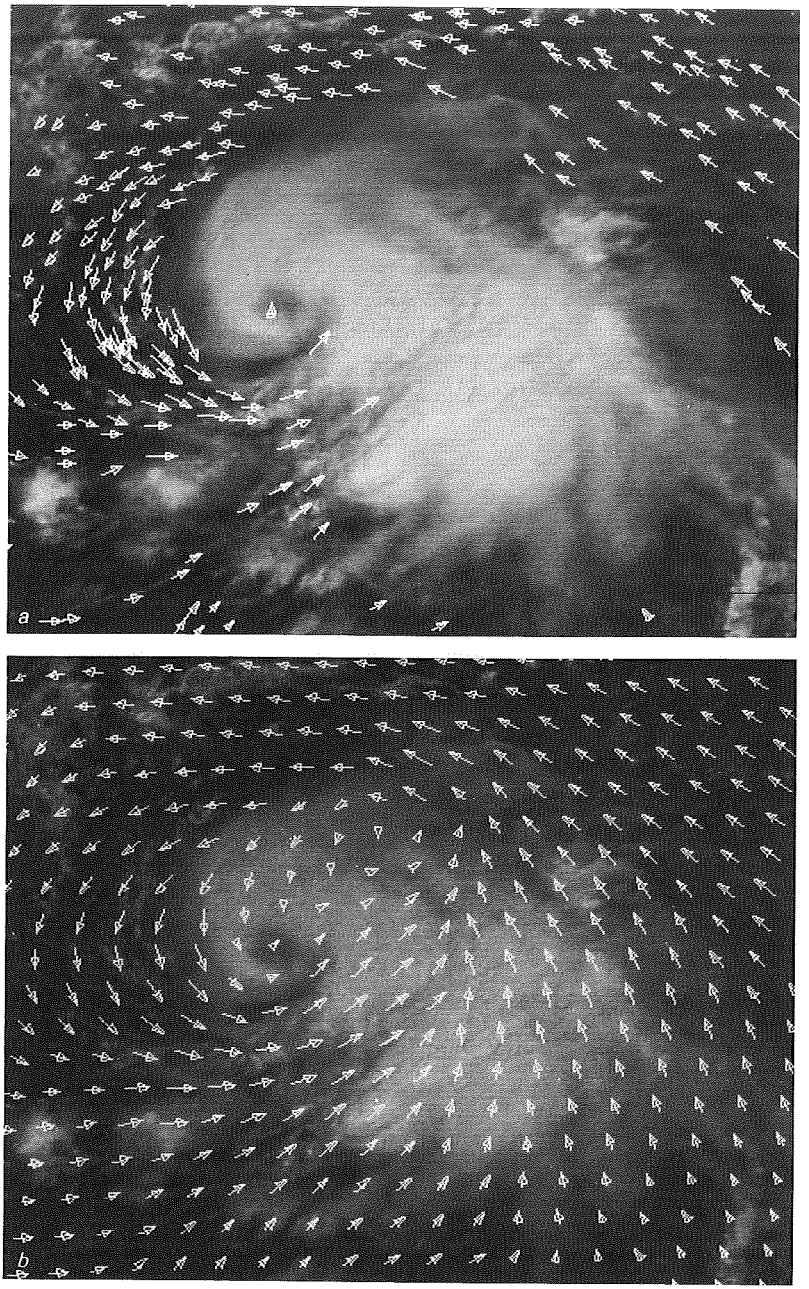


FIGURE 7.4—*Combination of visible image of tropical storm Anita with wind field. (a) Image combined with wind field derived by cloud tracking. (b) Image combined with interpolated wind field.*

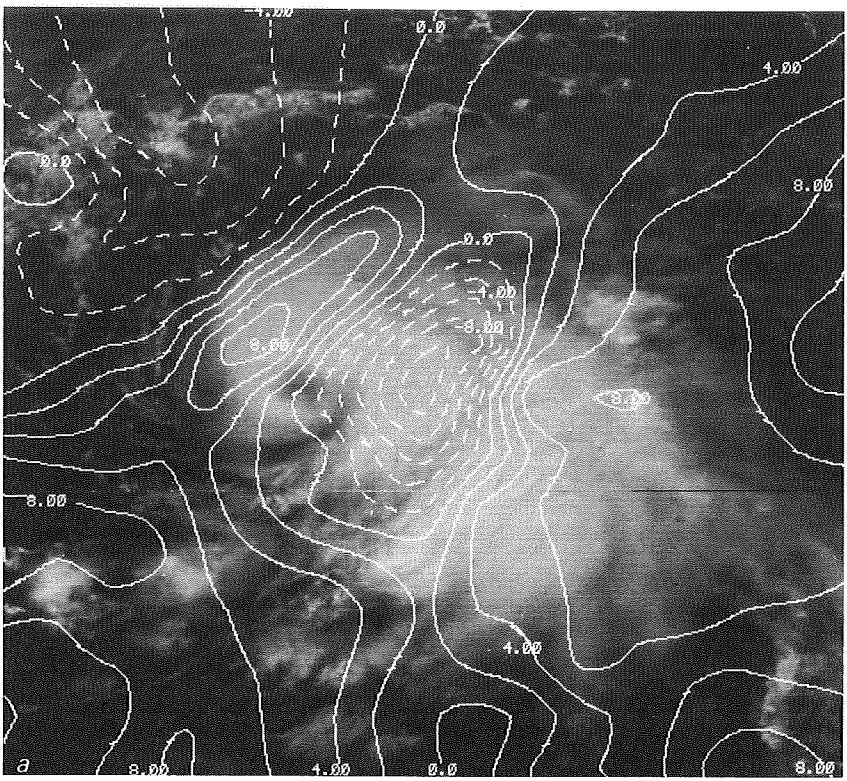


FIGURE 7.5a—*Radial wind velocity component in polar coordinates.*

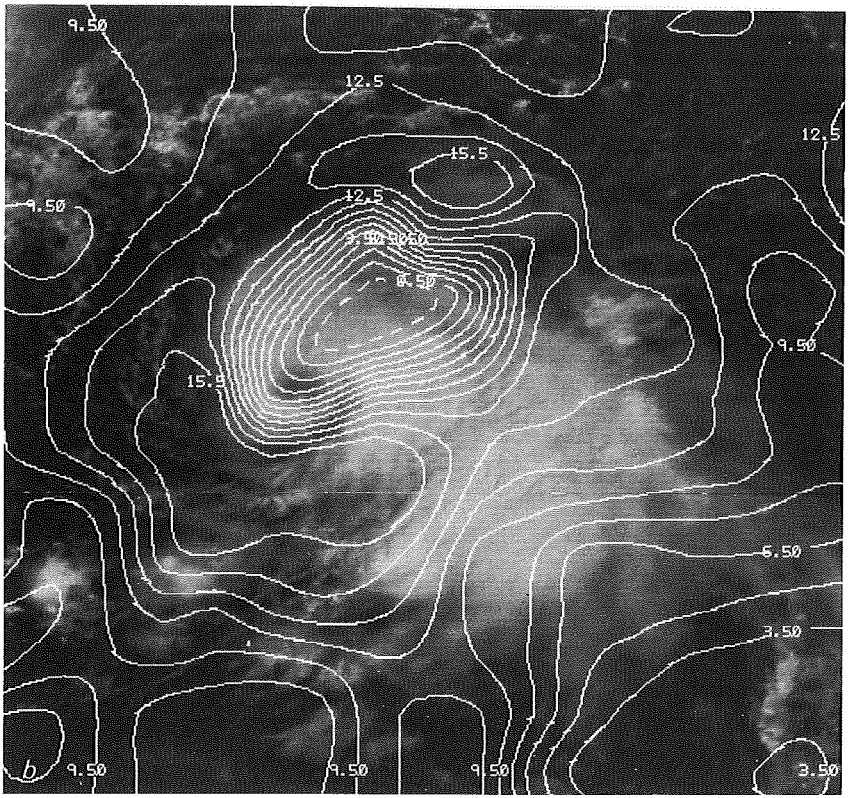


FIGURE 7.5b—*Tangential wind velocity component in polar coordinates.*

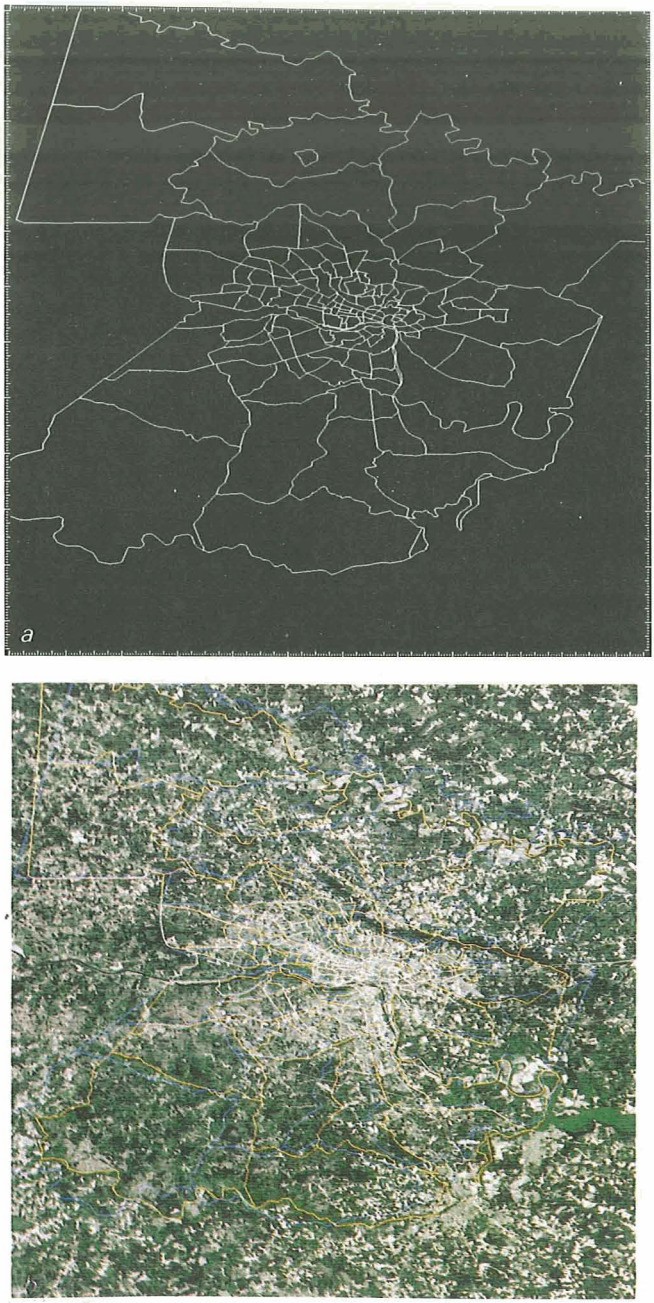


FIGURE 7.6—*Land-use mapping of Richmond, Va. area with IBIS. (a) Census tract boundaries. (b) Original and registered boundaries combined with Landsat MSS bands 4 and 5 image (scene 5340–14420).*

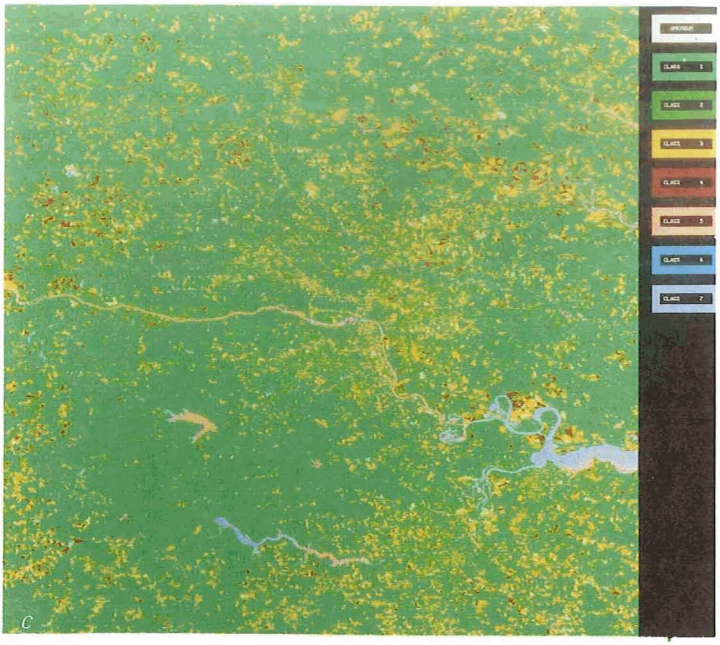


FIGURE 7.6—*Continued. (c) Unsupervised classification map. (d) Census tract map.*

sec. 8.4.) Seven different classes were distinguished. The third step involves the identification of each census tract in the boundary image with a unique color or gray value, generating a map as shown in figure 7.6*d*. The last major processing step is to combine the segmented image with the census tract map. Description of the census tract regions in terms of image properties is then simply a counting operation and report generation. A part of the report listing properties of the seven classes for the census tracts is shown in table 7.1.

7.4.3 Change Detection

An important application of remotely sensed images is the monitoring of changes on the Earth's surface caused by natural or manmade activities. Image segmentation techniques may be applied to detect temporal changes of specific objects or regions. Differencing of registered images (see sec. 4.5.2) shows all changes and causes errors in delineating the changed regions of interest. However, segmentation of a reference image into the regions of interest and background, and comparison of only these regions in a sequence of images permits detection of the specific changes.

In an application to assess the intensity and spatial distribution of forest insect damage in the Northeastern United States hardwood forests, a multispectral Landsat MSS image from one date is segmented by multidimensional thresholding into forest and nonforest regions [31]. All nonforest regions in the registered images from other dates are then eliminated, a step that permits an accurate detection of forest alterations without confusion caused by changes in agricultural areas and reduction of data volume. Figure 7.7 shows two Landsat MSS images obtained 1 year apart, the segmentation of one image into forest and nonforest regions, and the map displaying the forest areas changed by insect infestation.

TABLE 7.1—*Sample Land-Use Report for Richmond, Va.*

Point number	Tract	Region totals		Class 1		Class 2		Class 3		Class 4		Class 5		Class 6		Class 7	
		Number of pixels	Area (acres)	Area (acres)	Portion of total (percent)	Area (acres)	Portion of total (percent)	Area (acres)	Portion of total (percent)	Area (acres)	Portion of total (percent)	Area (acres)	Portion of total (percent)	Area (acres)	Portion of total (percent)	Area (acres)	Portion of total (percent)
51	2002.02	740	878	757	86.2	116	13.2	5	0.5	0	0.0	0	0.0	0	0.0	0	0.0
52	501.00	641	760	653	72.7	60	8.0	144	18.9	1	0.2	0	0.0	2	0.3	0	0.0
53	401.00	252	299	248	82.9	42	13.9	9	3.2	0	0.0	0	0.0	0	0.0	0	0.0
54	111.00	567	672	554	82.4	83	12.3	32	4.8	2	0.4	0	0.0	1	0.2	0	0.0
55	110.00	268	318	204	64.2	82	25.7	30	9.3	1	0.4	0	0.0	1	0.4	0	0.0
56	2001.07	1634	1938	1124	58.0	657	33.9	130	6.7	24	1.2	0	0.0	2	0.1	0	0.0
57	505.00	913	1083	830	76.7	180	16.6	32	3.0	36	3.3	0	0.0	2	0.2	2	0.2
58	502.00	579	687	543	79.1	65	9.5	76	11.1	0	0.0	0	0.0	2	0.3	0	0.0
59	2001.09	1044	1238	931	75.2	294	23.8	9	0.8	0	0.0	2	0.2	0	0.0	1	0.1
60	402.00	449	588	427	80.2	82	15.4	24	4.5	0	0.0	0	0.0	0	0.0	0	0.0
61	504.00	421	499	397	79.6	91	18.3	8	1.7	2	0.5	0	0.0	0	0.0	0	0.0
62	407.00	206	244	179	73.3	5	1.9	59	24.3	1	0.5	0	0.0	0	0.0	0	0.0
63	201.00	148	176	40	23.0	122	69.6	13	7.4	0	0.0	0	0.0	0	0.0	0	0.0
64	2012.02	865	1026	530	61.7	399	38.8	97	9.5	0	0.0	0	0.0	0	0.0	0	0.0
65	2001.08	5194	6160	4288	69.6	1413	22.9	114	1.8	219	3.6	117	1.9	0	0.0	9	0.2
66	406.00	93	110	88	80.6	0	0.0	21	19.4	0	0.0	0	0.0	0	0.0	0	0.0
67	301.00	154	163	66	52.6	21	11.7	64	35.1	0	0.0	0	0.0	1	0.8	0	0.0
68	204.00	372	441	63	14.2	235	53.2	116	26.3	27	6.2	0	0.0	0	0.0	0	0.0
69	2014.01	1412	1675	1006	60.1	400	23.9	250	14.9	1	0.1	0	0.0	18	1.1	0	0.0
70	405.00	189	224	151	67.2	43	19.0	30	13.2	0	0.0	0	0.0	1	0.5	0	0.0
71	202.00	180	213	79	37.2	84	39.4	20	9.4	28	13.3	0	0.0	1	0.6	0	0.0
72	409.00	211	250	231	92.4	5	1.9	14	5.7	0	0.0	0	0.0	0	0.0	0	0.0
73	404.00	140	166	100	60.0	18	10.7	49	29.3	0	0.0	0	0.0	0	0.0	0	0.0
74	302.00	169	200	74	36.7	55	27.2	63	31.4	0	0.0	0	0.0	9	4.7	0	0.0
75	203.00	141	167	57	34.0	71	42.6	15	9.2	24	14.2	0	0.0	0	0.0	0	0.0
76	----------	142	168	142	84.5	18	10.6	8	4.9	0	0.0	0	0.0	0	0.0	0	0.0
77	408.00	175	208	199	96.7	8	4.0	0	0.0	0	0.0	0	0.0	0	0.0	0	0.0
78	410.00	129	153	128	83.7	9	6.2	15	10.1	0	0.0	0	0.0	0	0.0	0	0.0
79	403.00	158	187	133	70.9	23	12.0	32	17.1	0	0.0	0	0.0	0	0.0	0	0.0
80	303.00	143	170	113	66.4	44	25.9	13	7.7	0	0.0	0	0.0	0	0.0	0	0.0
81	209.00	511	606	147	24.3	262	43.2	157	25.8	28	4.7	0	0.0	12	2.0	0	0.0

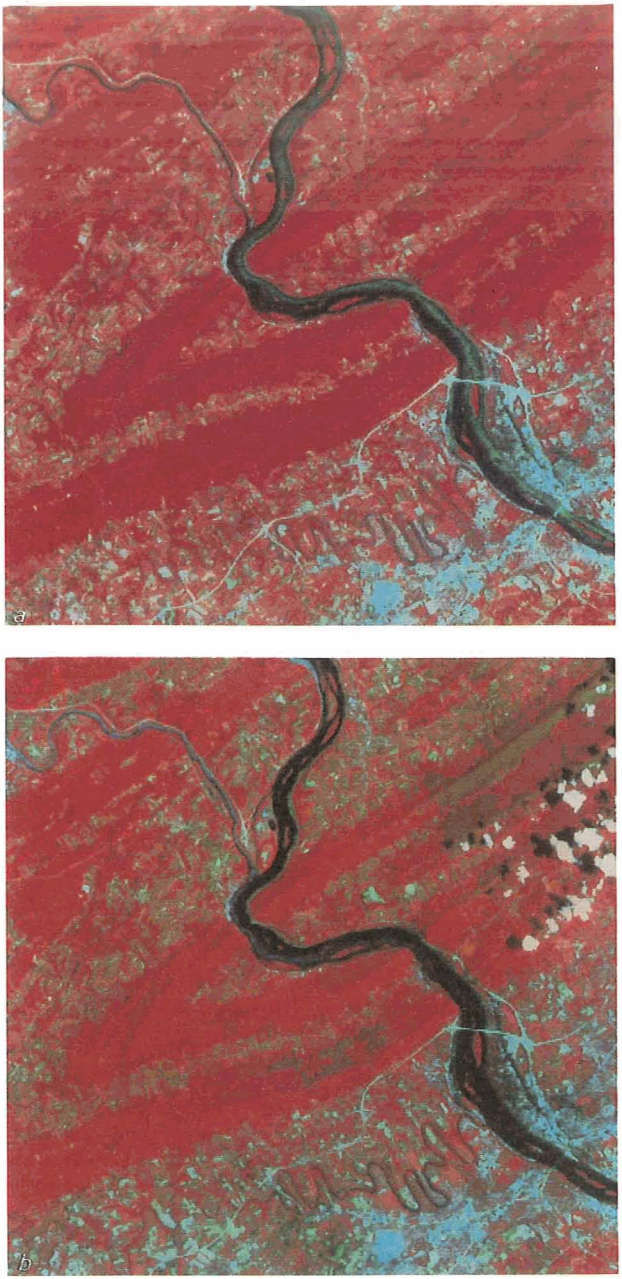


FIGURE 7.7—*Detection of forest cover alterations. (a) Landsat MSS image of Harrisburg, Pa. area taken July 19, 1976 (scene 544-15001). (b) Landsat MSS image with insect infestation taken June 27, 1977 (scene 2887–14520).*

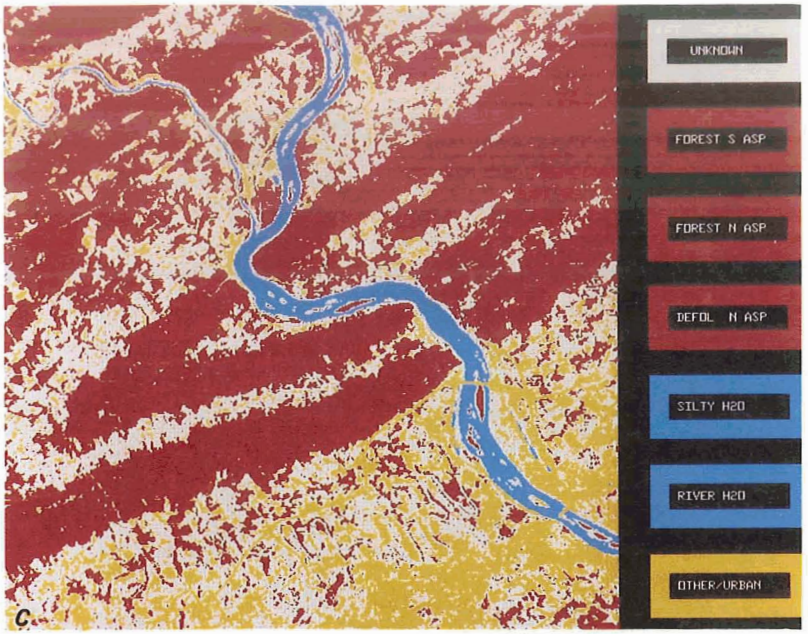


FIGURE 7.7—*Continued. (c) Segmented version of image in part a showing forest regions in red. (d) Change map showing forest regions in yellow and infested areas in blue.*

REFERENCES

[1] Rosenfeld, A.; and Kak, A. C.: Digital Picture Processing. Academic Press, New York, 1976.
[2] Prewitt, M. S.: Object Enhancement and Extraction, *in* Lipkin, B. S.; and Rosenfeld, A.: Picture Processing and Psychopictorics. Academic Press, New York and London, 1970, pp. 75–149.
[3] Duda, R. O.; and Hart, P. E.: Pattern Classification and Scene Analysis. Wiley-Interscience, New York and London, 1973.
[4] Rosenfeld, A.; and Thurston, M.: Edge and Curve Detection for Visual Scene Analysis, IEEE Trans. Comput., vol. C-20, 1971, pp. 562–569.
[5] Hueckel, M.: An Operator Which Locates Edges in Digital Pictures, J. Assoc. Comput. Mach., vol. 18, 1971, pp. 113–125.
[6] Nack, M. L.: Temporal Registration of Multispectral Digital Satellite Images Using Their Edge Images. AAS/AIAA Astrodynamics Specialist Conference, Nassau, Bahamas, July 1975.
[7] Fram, J. R.; and Deutsch, E. S.: On the Evaluation of Edge Detection Schemes and Their Comparison with Human Performance, IEEE Trans. Comput., vol. C-24, 1975, pp. 616–628.
[8] Ehrich, R. W.: Detection of Global Edges in Textured Images. Technical Report, ECE Dept., University of Massachusetts, Amherst, Mass., 1975.
[9] Frei, W.; and Chen, C.: Fast Boundary Detection: A Generalization and a New Algorithm, IEEE Trans. Comput., vol. C-26, 1977, pp. 988–998.
[10] Gupta, T. N.; and Wintz, P. A.: A Boundary Finding Algorithm and Its Applications, IEEE Trans. Circuits Syst., vol. CAS-22, 1975, pp. 351–362.
[11] Robertson, T. V.; Fu, K. S.; and Swain, P. H.: Multispectral Image Partitioning. LARS Information Note 071373, Purdue University, Lafayette, Ind., 1973.
[12] Zucker, S. W.; Rosenfeld, A.; and Davis, L. S.: Picture Segmentation by Texture Discrimination, IEEE Trans. Comput., vol. C-24, 1975, pp. 1228–1233.
[13] Hawkins, J. K.: Textural Properties for Pattern Recognition, *in* Lipkin, B. S.; and Rosenfeld, A.: Picture Processing and Psychopictorics. Academic Press, New York and London, 1970, pp. 347–370.
[14] Weszka, J. S.; Dyer, C. R.; and Rosenfeld, A.: A Comparative Study of Texture Measures for Terrain Classification, IEEE Trans. Systems, Man Cybernetics, vol. SMC-6, 1976, pp. 269–285.
[15] Haralick, R. M.; Shanmugam, K.; and Dirnstein, I.: Texture Features for Image Classification, IEEE Trans. Systems, Man Cybernetics, vol. SMC-3, 1973, pp. 610–621.
[16] Brayer, J. M.; and Fu, K. S.: Application of Web Grammar Model to an Earth Resources Satellite Picture. Proceedings of Third International Joint Conference on Pattern Recognition, Coronado, Calif., 1976.
[17] Pfaltz, J. L.; and Rosenfeld, A.: Web Grammars. Proceedings of First International Joint Conference on Artificial Intelligence, Washington, D.C., 1969.
[18] Miller, W. F.; and Shaw, A. C.: Linguistic Methods in Picture Processing—A Survey. Proceedings of Fall Joint Computer Conference, Thompson, Washington, D.C., 1968, pp. 279–290.
[19] Fu, K. S.: Syntactic Methods in Pattern Recognition. Academic Press, New York, 1974.
[20] Bracken, P. A.; Dalton, J. T.; Quann, J. J.; and Billingsley, J. B.: AOIPS—An Interactive Image Processing System. National Computer Conference Proceedings, AFIPS Press, 1978, pp. 159–171.
[21] Moik, J. G.: Smips/VICAR Image Processing System—Application Program Description. NASA TM 80255, 1979.
[22] Hubert, L. F.; and Whitney, L. F., Jr.: Wind Estimation from Geostationary Satellite Pictures, Mon. Weather Rev., vol. 99, 1971, pp. 665–672.
[23] Arking, A.; Lo, R. C.; and Rosenfeld, A.: A Fourier Approach to Cloud Motion Estimation, J. Appl. Meteorol., vol. 17, 1978, pp. 735–744.

[24] Smith, E. A.; and Phillips, D. R.: Automated Cloud Tracking Using Precisely Aligned Digital ATS Pictures, IEEE Trans. Comput., vol. C-21, 1972, pp. 715–729.

[25] Mottershead, C. T.; and Phillips, D. R.: Image Navigation for Geosynchronous Meteorological Satellites. Seventh Conference on Aerospace and Aeronautical Meteorology and Symposium on Remote Sensing from Satellites, American Meteorological Society, Melbourne, Fla., 1976, pp. 260–264.

[26] Leese, J. A.; Novak, C. S.; and Clark, B. B.: An Automated Technique for Obtaining Cloud Motion from Geosynchronous Satellite Data Using Cross-correlation, J. Appl. Meteorol., vol. 10, 1971, pp. 118–132.

[27] Billingsley, J.; Chen, J.; Mottershead, C.; Bellian, A.; and DeMott, T.: AOIPS Metpak—A Meteorological Data Processing System. Computer Sciences Corp. Report CSC/SD-77/6084, 1977.

[28] Rodgers, E.; Gentry, R. C.; Shenk, W.; and Oliver, V.: The Benefits of Using Short Interval Satellite Images to Derive Winds for Tropical Cyclones, Mon. Weather Rev., vol. 107, May 1979.

[29] Bryant, N. A.; and Zobrist, A. L.; IBIS: A Geographic Information System Based on Digital Image Processing and Image Raster Datatype. Proceedings of Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Lafayette, Ind., 1976, p. 1A-1 to 1A-7.

[30] Bryant, N. A.: Integration of Socioeconomic Data and Remotely Sensed Imagery for Land Use Applications. Proceedings of Caltech/JPL Conference on Image Processing Technology, Data Sources and Software for Commercial and Scientific Applications, California Institute of Technology, Pasadena, Calif., Nov. 1976, pp. 9-1–9-8.

[31] Williams, D. L.; and Stouffer, M. L.: Monitoring Gvpsv Moth Defoliation Via Landsat Image Differencing. Symposium on Remote Sensing for Vegetation Damage Assessment, American Society of Photogrammetry, 1978, pp. 221–229.

8. Image Classification

8.1 Introduction

An important segmentation method for multi-images is classification, whereby objects (points or regions) of an image are assigned to one of a prespecified set of classes. The description is simply the name of the class to which the object belongs. A multi-image is represented by a set of property values measured or computed for each component. A property may be the gray level, a texture measure, the coefficient of an orthogonal image transformation (e.g., Fourier, Karhunen–Loéve transformation), or the description of size and shape of a region in the image. The set of property values for a given point is called a pattern.

The property values are combined into a P-dimensional vector $\mathbf{f}$, which can be represented as a point in pattern space. The basic condition for classification is that the representative patterns of a class form compact regions or clusters in pattern space, i.e., that the pattern vectors are not randomly distributed. The assumption is then made that each pattern $\mathbf{f}$ belongs to one and only one of K classes. For remotely sensed image data, this assumption is justified, because most materials reflect or emit a unique spectrum of electromagnetic energy. (For example, see fig. 8.1.)

Because of the variations of object characteristics and noise, remotely sensed images may be regarded as samples of random processes. (See sec. 2.2.2). Thus, image properties or patterns are random variables. The variations and clustering of pattern vectors with the spectral characteristics shown in figure 8.1 are illustrated in figure 8.2.

Given a set of patterns for an image, statistical decision theory or geometric techniques may be used to decide to which class a pattern should be assigned. The set of decision rules is called a classifier. Let the set of all patterns be S, where

$$S=\{{}^1\mathbf{f}, {}^2\mathbf{f}, \ldots\} \tag{8.1}$$

Formally, the classes are obtained by partitioning the set S into K subsets S_k such that

$$S_k \cap S_j = \varnothing \qquad k \neq j$$

$$\bigcup_{k=1}^{K} S_k = S \tag{8.2}$$

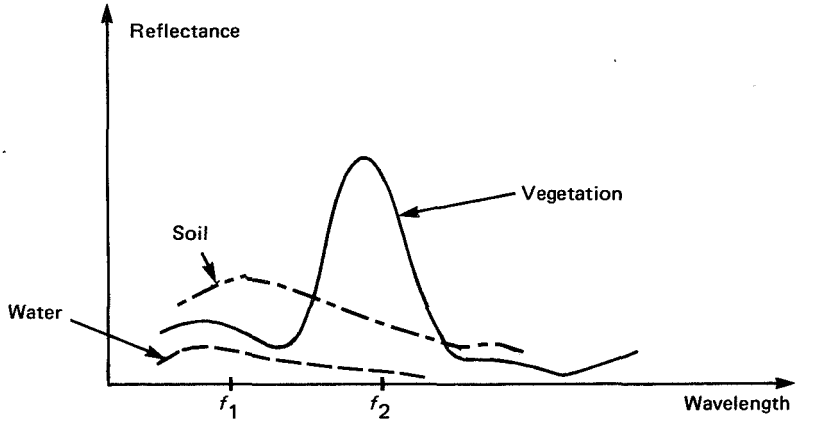


FIGURE 8.1—*Spectral characteristics of different object types.*

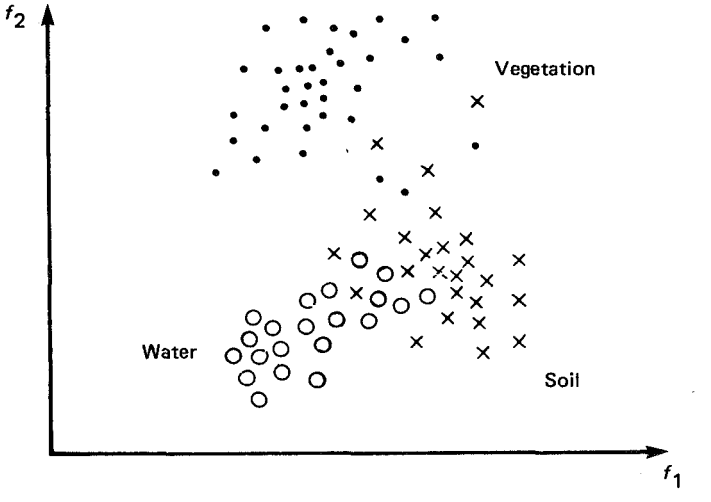


FIGURE 8.2—*Clusters in pattern space.*

The design of a classifier requires some information about the set S. In general this information is incomplete, and only a subset s is known, where

$$s = \{{}^1\mathbf{f}, \ldots, {}^M\mathbf{f}\} \subset S \tag{8.3}$$

The set s is called the training set. It is used to obtain information about

the classes S_k and to derive the class boundaries. Depending on the available knowledge, the following cases may be distinguished:

1. The training set s is available, and a partition into K subsets s_k is known, such that

$$s_k \subset S_k \tag{8.4}$$

Thus, the class membership of each pattern in the training set is known. It is assumed that each subset s_k contains M_k training patterns $\mathbf{f}_k$. This case is known as supervised classification.

2. A training set s is available. However, the partition into subsets s_k is unknown. This case is known as unsupervised classification. The problem is further complicated if the number of classes K is also unknown.

The classification problem is to find a decision rule (a classifier) that partitions an image into K disjoint class regions according to knowledge about a limited set of patterns. The classifier should be effective for patterns not in the training set, and it should be efficient with respect to execution time. Because of the random variations, a given pattern may belong to more than one class. To reduce ambiguous decisions and to assign a pattern to only one class, an additional reject class S_0 is often introduced. All patterns with dubious class membership are assigned to S_0.

Image classification consists of the following steps (fig. 8.3): (1) Preprocessing, (2) training set selection and determination of class characteristics, (3) feature selection, and (4) classification.

The digitized images are preprocessed to correct for radiometric and geometric errors and enhanced to facilitate the selection of a training set. This training set s is used to determine the class characteristics with a supervised technique if the partition of s into K classes is known, or with an unsupervised technique if no partition of s is known. Feature selection determines a set of image properties that best describe and discriminate object categories. These properties are called features. Often, the original measurements, i.e., the gray values of a multi-image, are used as features. Finally, classification of the image patterns based on the selected features is performed.

It is assumed that the training set s contains M patterns ${}^j\mathbf{f}$:

$$s = \{{}^j\mathbf{f},\ j = 1, \ldots, M\} \tag{8.5}$$

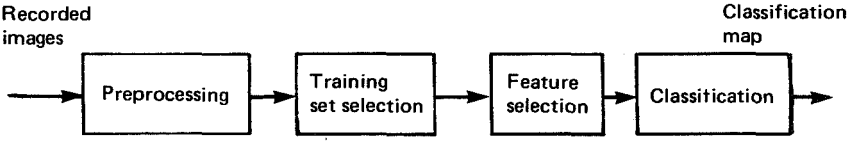


FIGURE 8.3—*Image classification.*

For supervised classification, the training patterns for class S_k are denoted by ${}^j\mathbf{f}_k$. Thus, s_k is given by

$$s_k=\{{}^j\mathbf{f}_k,\ j=1,\ldots,M_k\} \qquad k=1,\ldots,K \tag{8.6}$$

where it is assumed that M_k training patterns are given for class S_k and that the number of classes is K. The features for a given point or region in an image will be represented by an N-dimensional feature vector $\mathbf{z}$.

The structure of statistical classifiers is determined primarily by the probability density function $p(\mathbf{z})$ for the feature vectors. Most important is the multivariate normal (Gaussian) density, given by

$$p(\mathbf{z})=\frac{1}{(2\pi)^{N/2}\,|\mathbf{C}|^{1/2}}e^{-\frac{1}{2}(\mathbf{z}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{z}-\mathbf{m})} \tag{8.7}$$

where $\mathbf{z}$ is an N-dimensional feature or pattern vector, $\mathbf{m}$ is the N-dimensional mean vector, $\mathbf{C}=(\sigma_{ij})$ is the N by N covariance matrix, and $|\mathbf{C}|$ is the determinant of $\mathbf{C}$. Mean vector and covariance are defined by

$$\mathbf{m}=E\{\mathbf{z}\} \tag{8.8}$$

and

$$\mathbf{C}=E\{(\mathbf{z}-\mathbf{m})^T(\mathbf{z}-\mathbf{m})\} \tag{8.9}$$

where the expected value of a vector or matrix is found by taking the expected values of its components. The covariance matrix is always symmetric. The diagonal element σ_{ii} is the variance of z_i, and the off-diagonal element σ_{ij} is the covariance of z_i and z_j. If z_i and z_j are statistically independent, $\sigma_{ij}=0$. The matrix $\mathbf{C}$ is positive definite; so $|\mathbf{C}|$ is strictly positive. Cases where $|\mathbf{C}|=0$ would occur (e.g., when one component of $\mathbf{z}$ has zero variance or when two components are identical) are excluded.

The multivariate normal density is completely specified by the mean vector and the covariance matrix. Knowledge of the covariance matrix allows calculation of the dispersion of the data in any direction. Normally distributed patterns form a single cluster (fig. 8.4). The center of the cluster is determined by the mean vector, and the shape of the cluster is determined by the covariance matrix. The loci of points of constant density, as given by equation (8.7), are hyperellipsoids for which the quadratic form is constant:

$$d=(\mathbf{z}-\mathbf{m})^T\mathbf{C}^{-1}(\mathbf{z}-\mathbf{m}) \tag{8.10}$$

The quantity d is called the Mahalanobis distance from $\mathbf{z}$ to $\mathbf{m}$. The principal axes of the hyperellipsoids are given by the eigenvectors of $\mathbf{C}$, and the eigenvalues determine the length of the axes. The volume of the hyperellipsoid corresponding to a Mahalanobis distance d is given by

$$V=V_N\,|\mathbf{C}|^{\frac{1}{2}}\,d^N \tag{8.11}$$

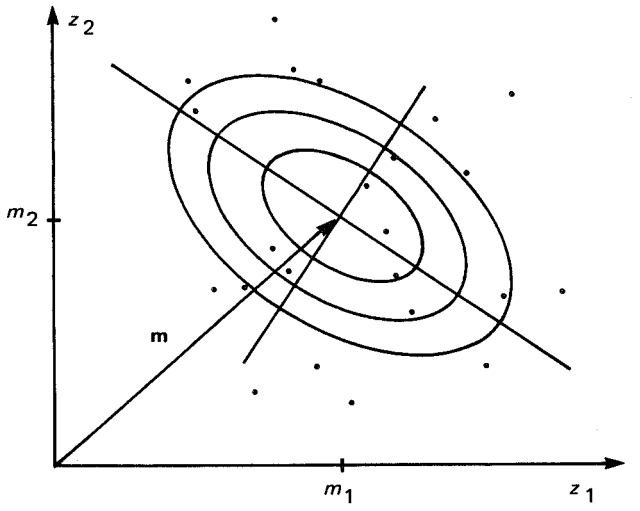


FIGURE 8.4—*Bivariate normal distribution.*

where V_N is the volume of an N-dimensional unit hypersphere:

$$V_N = \begin{cases} \dfrac{\pi^{N/2}}{(N/2)!} & N \text{ even} \\ \dfrac{2^N \pi^{(N-1)/2} \left(\dfrac{N-1}{2}\right)!}{N!} & N \text{ odd} \end{cases} \tag{8.12}$$

Thus, for a given pattern dimensionality N, the scatter of the patterns varies directly with $|\mathbf{C}|^{1/2}$.

In general $\mathbf{m}$ and $\mathbf{C}$ are not known. They can be estimated from the M training patterns ${}^j\mathbf{z}$ by

$$\mathbf{m} = \frac{1}{M} \sum_{j=1}^{M} {}^j\mathbf{z} \tag{8.13}$$

and

$$\mathbf{C} = \frac{1}{M-1} \sum_{j=1}^{M} ({}^j\mathbf{z} - \mathbf{m})({}^j\mathbf{z} - \mathbf{m})^T \tag{8.14}$$

8.2 Feature Selection

The representation of patterns in terms of measured image properties often does not lead to efficient classification schemes, because the classes may be difficult to separate or the number of measurements is large, or both. One reason is that sensors are in general defined by other than pattern classification specifications. There may be linear or nonlinear

combinations of the measurements that afford a better separation of the classes. The concept of feature selection is used to determine those measurements that are most effective in classification. If the dimensionality P of measurement space is large (e.g., a scanner with 12 or more channels), classification algorithms cannot be efficiently implemented in measurement space, and classification with even simple algorithms becomes very time consuming on digital computers. It is therefore desirable to reduce the dimensionality of the space in which classification algorithms must be computed. Therefore, a feature space of dimensionality $N<P$ is introduced.

Features are sets of combined or selected measurements derived from the originally measured image properties. The goal is to find a set of features that has lower dimensionality than the original measurements and optimizes classifier performance. The features for a given spatial location (x, y) are represented by an N-dimensional feature vector $\mathbf{z}=(z_1, \ldots, z_N)^T$. Thus, feature selection is a mapping of the set of P-dimensional patterns $\{\mathbf{f}\}$ into the set of N-dimensional feature vectors $\{\mathbf{z}\}$. There is no general, theoretically justified method for feature selection. Many proposed techniques are based on heuristic considerations, such as intuition and experience. For example, in an interactive image analysis and recognition system the analyst with his previously acquired knowledge is part of the selection process [1]. In this context feature selection is more than just the transformation of the measurement space into a form that can simplify the pattern classification procedure. It provides for inserting a priori knowledge that reduces the number of required pattern samples to achieve a specified performance. A more mathematically founded technique is the expansion of a pattern into an orthonominal series (e.g., Fourier, Karhunen–Loéve expansion), where the expansion coefficients are used as features. Here new features with, it is hoped, lower dimensionality and better class discrimination are obtained by a linear transformation of the patterns [2, 3].

Another approach is the evaluation of the quality of pattern components and selection of a subset as features. For statistically independent pattern components, distance measures between the probability densities characterizing the pattern classes may be used. Two such distance measures that have been widely used are divergence and Battacharyya distance [4].

8.2.1 Orthogonal Transforms

The coefficients of expansion of the patterns $\mathbf{f}$ into a complete set of orthonormal matrices may be used as components of a feature vector $\mathbf{z}$. The feature vectors are obtained by a linear transformation:

$$\mathbf{z}=\mathbf{T}\mathbf{f} \tag{8.15}$$

where **T** is a unitary matrix whose rows are the basis vectors of the expansion. (See sec. 2.6.1.4.)

Expansions that are independent of the image patterns are the Fourier and Hadamard transforms. The orthonormal matrices are the harmonic and the Hadamard matrices. The expansion may also be adapted to the image characteristics. In this case the image patterns are considered random variables, and a training set is required to compute statistical characteristics and orthogonal matrices according to a given criterion. A criterion that minimizes the error of approximation of the patterns by the features without class distinction leads to the Karhunen–Loéve (K–L) transform. (See sec. 2.6.1.4.) The K–L transform determines the orthonormal vector system that provides the best approximation of the original patterns in a mean-square-error sense. The coefficients are ordered such that the patterns are represented by the fewest possible number of features. Furthermore, the obtained features are uncorrelated. In classification, however, the interest is in discrimination between patterns from different classes, not in accurate representation. Feature vectors should emphasize differences between classes. Therefore, other criteria that maximize the distance between feature vectors from different classes are used.

In section 2.6.1.4 it was shown that the basis vectors of the optimal expansion are obtained as eigenvectors of a symmetric, positive definite kernel matrix **R** which will now be denoted by **Q**. This matrix is computed from a training set. In section 8.1 two cases were distinguished: (1) a partition of the training set s into K classes s_k is known, and (2) the class membership of the training patterns is not known.

In the first case, the error criterion, equation (2.127), has to be modified to account for the class membership of the patterns. A mean-square approximation error ϵ is defined as

$$\epsilon = \sum_{k=1}^{K} P(S_k) E\left\{ || \mathbf{f}_k - \sum_{n=1}^{N} z_{kn} \mathbf{t}_n ||^2 \right\} \tag{8.16}$$

where $P(S_k)$ is the a priori class probability, $\mathbf{f}_k$ is a P-dimensional training pattern from class s_k, $\{\mathbf{t}_n\}$ is a set of orthogonal vectors, and the expansion coefficients z_{kn} are used as elements of the new class feature vectors $\mathbf{z}_k$. The best approximation of the original patterns in a mean-square-error sense is given by determining $\{\mathbf{t}_n\}$ such that equation (8.16) is minimized. The new features are given by

$$z_{kn} = \mathbf{t}_n{}^T \mathbf{f}_k \qquad n = 1, \ldots, N \tag{8.17}$$

Proceeding as in section 2.6.1.4, the error, as given in equation (8.16), becomes

$$\epsilon = \sum_{n=N+1}^{P} \mathbf{t}_n{}^T \left(\sum_{k=1}^{K} P(S_k) E\{\mathbf{f}_k \mathbf{f}_k{}^T\} \right) \mathbf{t}_n \tag{8.18}$$

Thus, the kernel matrix $\mathbf{Q}$ is the overall correlation matrix

$$\mathbf{R}=\sum_{k=1}^{K} P(S_k)E\{\mathbf{f}_k\mathbf{f}_k{}^T\} \tag{8.19}$$

which is a weighted average of the class correlation matrices $\mathbf{R}_k$, where

$$\mathbf{R}_k=E\{\mathbf{f}_k\mathbf{f}_k{}^T\} \tag{8.20}$$

For patterns with nonzero means, the class covariance matrices $\mathbf{C}_k$ are used:

$$\mathbf{C}_k=E\{(\mathbf{f}_k-\mathbf{m}_k)(\mathbf{f}_k-\mathbf{m}_k)^T\} \tag{8.21}$$

If the a priori class probabilities $P(S_k)$ are all equal, the total covariance matrix $\mathbf{C}$ becomes

$$\mathbf{C}=\frac{1}{K}\sum_{k=1}^{K}\mathbf{C}_k \tag{8.22}$$

Minimization of equation (8.18) (see sec. 2.6.1.4) leads to the following eigenvalue problem:

$$\mathbf{C}\mathbf{t}_n=\lambda_n\mathbf{t}_n \tag{2.130}$$

The first N of the P orthogonal eigenvectors of $\mathbf{C}$ are the rows of $\mathbf{T}$:

$$\mathbf{T}=\begin{pmatrix}\mathbf{t}_1{}^T\\ \cdot\\ \cdot\\ \cdot\\ \mathbf{t}_N{}^T\end{pmatrix} \tag{2.132}$$

Mean vector $\mathbf{n}_k$ and covariance matrix $\mathbf{D}_k$ of the class feature vectors $\mathbf{z}_k$ are given by

$$\mathbf{n}_k=E\{\mathbf{z}_k\}=\mathbf{T}\mathbf{m}_k \tag{8.23}$$

and

$$\left.\begin{aligned}\mathbf{D}_k&=E\{(\mathbf{z}_k-\mathbf{n}_k)(\mathbf{z}_k-\mathbf{n}_k)^T\}\\ &=\mathbf{T}\mathbf{C}_k\mathbf{T}^T=\operatorname{diag}(\lambda_{kn})=\operatorname{diag}(\sigma_{kn}{}^2)\end{aligned}\right\} \tag{8.24}$$

Thus $\mathbf{D}_k$ is a diagonal matrix whose elements are the eigenvalues of $\mathbf{C}_k$ and the variances of the new features as well. If the patterns $\mathbf{f}_k$ for class S_k are normally distributed with the probability density

$$p(\mathbf{f}_k \mid S_k)=\frac{1}{(2\pi)^{P/2}\,|\mathbf{C}_k|^{1/2}}e^{-\frac{1}{2}(\mathbf{f}_k-\mathbf{m}_k)^T\mathbf{C}_k^{-1}(\mathbf{f}_k-\mathbf{m}_k)} \tag{8.25}$$

then the features $\mathbf{z}_k$ obtained by equation (8.15) are normally distributed with the following density function:

$$p(\mathbf{z}_k \mid S_k) = \frac{1}{(2\pi)^{N/2} |\mathbf{D}_k|^{1/2}} e^{-\frac{1}{2}(\mathbf{z}_k - \mathbf{n}_k)^T \mathbf{D}_k^{-1} (\mathbf{z}_k - \mathbf{n}_k)} \tag{8.26}$$

because of equations (8.23) and (8.24). Expanding the exponent yields

$$(\mathbf{z}_k - \mathbf{n}_k)^T \mathbf{D}_k^{-1} (\mathbf{z}_k - \mathbf{n}_k) = \sum_{i=1}^{N} \frac{(z_{ki} - n_{ki})^2}{\lambda_i} \tag{8.27}$$

It then becomes evident that the contours of constant probability density are hyperellipsoids with centers at $\mathbf{n}_k$. The direction of the principal axis is along the eigenvectors of the covariance matrix and the diameters of a hyperellipsoid are proportional to the square roots of the corresponding eigenvalues or variances:

$$\lambda_{kn} = \sigma_{kn}.^2$$

The expansion based on equation (8.16) is known as generalized K–L transform [5]. Geometrically, the transformations (8.15) and (2.129), are a rotation of the pattern space. Figure 8.5 illustrates such a two-dimensional rotation for two classes. Feature selection by means of the generalized K–L transform is made irrespective of the mean vectors of the individual classes. To retain the means and, therefore, the spatial distribution of individual classes, the total covariance matrix $\mathbf{C}$ may be defined as

$$\mathbf{C} = \sum_{k=1}^{K} P(S_k) E\{(\mathbf{f}_k - \mathbf{m})(\mathbf{f}_k - \mathbf{m})^T\} \tag{8.28}$$

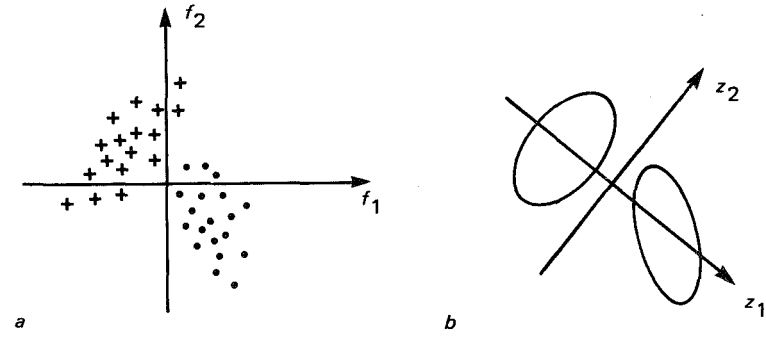


FIGURE 8.5—*Rotation of pattern space. (a) Pattern space. (b) Rotated pattern space.*

where

$$\mathbf{m}=E\{\mathbf{f}\} \tag{8.29}$$

The vector $\mathbf{m}$ is the total mean of all patterns in the training set s. Here the transform matrix $\mathbf{T}$ is a function of both the variances and the means of each class.

For classification the interest is in transforms that emphasize the dissimilarity between patterns of different classes rather than provide fidelity of representation. One possibility is to determine the linear transformation (8.15), such that the mean square interclass distance d is maximized, where

$$d=\frac{2}{K(K-1)}\sum_{k=2}^{K}\sum_{l=1}^{k-1} d_{kl}^{2} \tag{8.30}$$

The quantity d_{kl} is the mean distance between two feature vectors from different classes k and l. For each vector in class k the distance to all vectors in classes $l=1$ to $k-1$ is computed, and this computation is performed for $k=2$ to K. The mean distance between feature vectors of different classes may be defined as the Euclidean distance:

$$d_{kl}^{2}=E\{(\mathbf{z}_k-\mathbf{z}_l)^T(\mathbf{z}_k-\mathbf{z}_l)\} \tag{8.31}$$

With equation (8.17), the mean square distance d is obtained:

$$d=\sum_{n=1}^{N}\mathbf{t}_n^T\left[\frac{1}{K}\sum_{k=1}^{K}\mathbf{R}_k-\frac{1}{K(K-1)}\sum_{k=2}^{K}\sum_{l=1}^{k-1}(\mathbf{m}_k\mathbf{m}_l^T+\mathbf{m}_l\mathbf{m}_k^T)\right]\mathbf{t}_n \tag{8.32}$$

where $\mathbf{R}_k$ is the class correlation matrix given in equation (8.20) and $\mathbf{m}_k$, $\mathbf{m}_l$ are class means. Thus the distance d can be written as

$$d=\sum_{n=1}^{N}\mathbf{t}_n^T\mathbf{C}\mathbf{t}_n \tag{8.33}$$

with the kernel matrix $\mathbf{C}$, where

$$\mathbf{C}=\frac{1}{K}\sum_{k=1}^{K}\mathbf{R}_k-\frac{1}{K(K-1)}\sum_{k=2}^{K}\sum_{l=1}^{k-1}(\mathbf{m}_k\mathbf{m}_l^T+\mathbf{m}_l\mathbf{m}_k^T) \tag{8.34}$$

As in section 2.6.1.4 the vectors $\mathbf{t}_n$ that maximize d in equation (8.33) are obtained as eigenvectors of $\mathbf{C}$:

$$\mathbf{C}\mathbf{t}_n=\lambda_n\mathbf{t}_n \tag{2.130}$$

The vectors $\mathbf{t}_n$ are combined into the transform matrix $\mathbf{T}$ as in equation (2.133).

In the second case no partition of the training set s into classes is known. The optimal expansion is the K–L transform defined in section

2.6.1.4. The covariance matrix $\mathbf{C}$ is computed from the M training patterns $\mathbf{f}$ in s without class distinction as

$$\mathbf{C}=E\{(\mathbf{f}-\mathbf{m})(\mathbf{f}-\mathbf{m})^T\} \tag{8.35}$$

and $\mathbf{m}$ is the mean vector of the patterns in s.

8.2.2 Evaluation of Given Features

In the previous section new features were obtained by a linear transformation of the patterns. Class separability may be increased by proper choice of the criterion used to derive the transformation. The resulting P components are ordered according to the magnitude of their variance. Therefore, dimensionality reduction may be achieved by selecting the first $N<P$ components as features for classification. Another approach is to evaluate the P given pattern components and to select a subset of $N<P$ features that yields the smallest classification error. Ideally, this problem could be solved by computing the probability of correct classification associated with each N-feature subset and then selecting the one giving the best performance. However, it is generally not feasible to perform the required computations. The number of subsets of features that must be examined is

$$\binom{P}{N}=\frac{P!}{N!(P-N)!} \tag{8.36}$$

For example, to select the best four of eight available features requires 70 computations of the error probabilities. Therefore, alternative methods must be found for feature selection.

The distances between the class probability distributions may be used for feature evaluation. Intuitively, a feature for which the distance between class means is large and the sum of variances is small affords good class separation. With the assumption that the original features are statistically independent, the evaluation is simplified. Each of the P features is evaluated independently and the N best features are selected. For two classes S_j and S_k with mean vectors $\mathbf{m}_j$, $\mathbf{m}_k$ and variances σ_j^2 and σ_k^2, a quality measure for feature z_n, $n=1, \ldots, N$ may be defined as

$$G_n=\frac{(m_{jn}-m_{kn})^2}{\sigma_{jn}^2+\sigma_{kn}^2} \tag{8.37}$$

Obviously $0\leq G_n\leq\infty$. A large value for G_n indicates that feature z_n is useful for separating classes S_j and S_k.

Figure 8.6 shows examples for the applicability of the measure given in equation (8.37). In figure 8.6*a* the feature z_n is sufficient to separate the two classes S_j and S_k, but the large overlapping area in figure 8.6*b* requires additional features. The limits of the measure given in equation

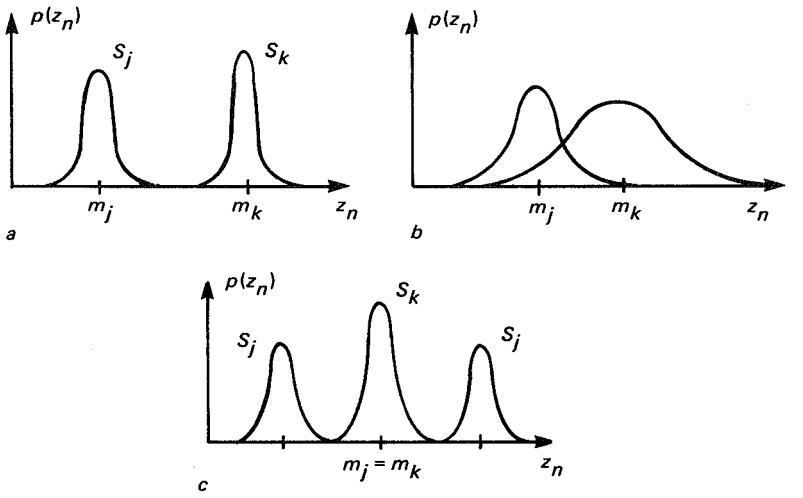


FIGURE 8.6—*Separability of features.*

(8.37) become obvious in figure 8.6*c*, where the distribution of class S_j has two maxima for feature z_n. Although complete separation is possible (there is no overlap of the distributions), $G_n=0$ gives no indication of the quality of feature z_n. Thus, if the class probability densities are not normal, mean and variance are not sufficient to evaluate the separability. Therefore, distance measures that are dependent on the distribution must be used.

One measure of the distance between classes is known as divergence [4]. Divergence between two classes S_j and S_k is defined as

$$D(S_j, S_k) = \int \ln \frac{p(\mathbf{z} \mid S_j)}{p(\mathbf{z} \mid S_k)} \; p(\mathbf{z} \mid S_j) - p(\mathbf{z} \mid S_k) d\mathbf{z} \qquad (8.38)$$

where $p(\mathbf{z} \mid S_k)$ is the probability density distribution of $\mathbf{z}$ for class S_k. Divergence is a measure of the dissimilarity of two distributions and thus provides an indirect measure of the ability of the classifier to discriminate successfully between them. Computation of this measure for groups of N of the available features provides a basis for selecting an optimal set of N features. The subset of N features for which D is maximum is best suited for separation of the two classes S_j and S_k. In the case of normal distributions with mean $\mathbf{m}_k$ and covariance matrix $\mathbf{C}_k$, the divergence becomes

$$\begin{aligned} D(S_j, S_k) = &\frac{1}{2} \operatorname{tr} [(\mathbf{C}_j - \mathbf{C}_k)(\mathbf{C}_j^{-1} - \mathbf{C}_k^{-1})] \\ &+ \frac{1}{2} \operatorname{tr} [(\mathbf{C}_j^{-1} - \mathbf{C}_k^{-1})(\mathbf{m}_j - \mathbf{m}_k)(\mathbf{m}_j - \mathbf{m}_k)^T] \end{aligned} \qquad (8.39)$$

where tr [**C**] is the trace of the matrix **C**.

Divergence is defined for two classes. An extension to K classes is possible by computing the average divergence D_A over all pairs of classes and selecting the subset of N features for which the average divergence is maximum; that is, by maximizing D_A with respect to all N-dimensional features, where

$$D_A(\mathbf{z}) = \frac{2}{K(K-1)} \sum_{k=1}^{K-1} \sum_{j=k+1}^{K} D(S_j, S_k) \tag{8.40}$$

This strategy, although reasonable, is not optimal. For instance, a large single pairwise divergence term in equation (8.40) could significantly bias the average divergence. So in the process of ranking feature combinations by D_A, it is useful to examine each of the pairwise divergences as well.

The behavior of the pairwise divergence $D(S_j, S_k)$ with respect to the probability of correct classification P_c is inconsistent. As the separability of a pair of classes increases, $D(S_j, S_k)$ also increases without limit, whereas P_c saturates at 100 percent. A modified form of the divergence, referred to as the transformed divergence D_T, has a behavior more like probability of correct classification [6, 7]:

$$D_T(S_j, S_k) = 100(1 - e^{-D(S_j, S_k)/8}) \tag{8.41}$$

The saturating behavior of D_T reduces the effects of widely separated classes when taking the average over all pairwise separations. The average divergence based on transformed divergence has been found a much more reliable criterion for feature selection than the average divergence based on ordinary divergence.

As an example, consider feature selection for the classification of forest types in a Landsat image. Figure 8.7 shows a Landsat MSS false-color image of an area in North Carolina, obtained February 26, 1974. Polygonal training areas representing the prototypes for seven classes are outlined in white. The class names and the number M_k of training vectors for each class are listed in table 8.1. The total covariance matrices computed from the training vectors with equations (8.22), (8.28), and (8.34), the eigenvalues and eigenvectors of the covariance matrices and the percentage variance in the principal components are shown in table 8.2.

The pairwise-transformed divergences for the $P=4$ features and the $K=7$ classes computed from the training data for the multispectral image in figure 8.6 are listed in the left column of table 8.3. Classes 1 and 3 (closed canopy and partial close) can not be reliably separated with the available features. To improve separability of the given classes, additional features derived from other measurements are required. These measurements may include other spectral bands, images obtained at different times or by other sensors and ground measurements. The pairwise transformed divergences for eight features obtained by addition of a second Landsat

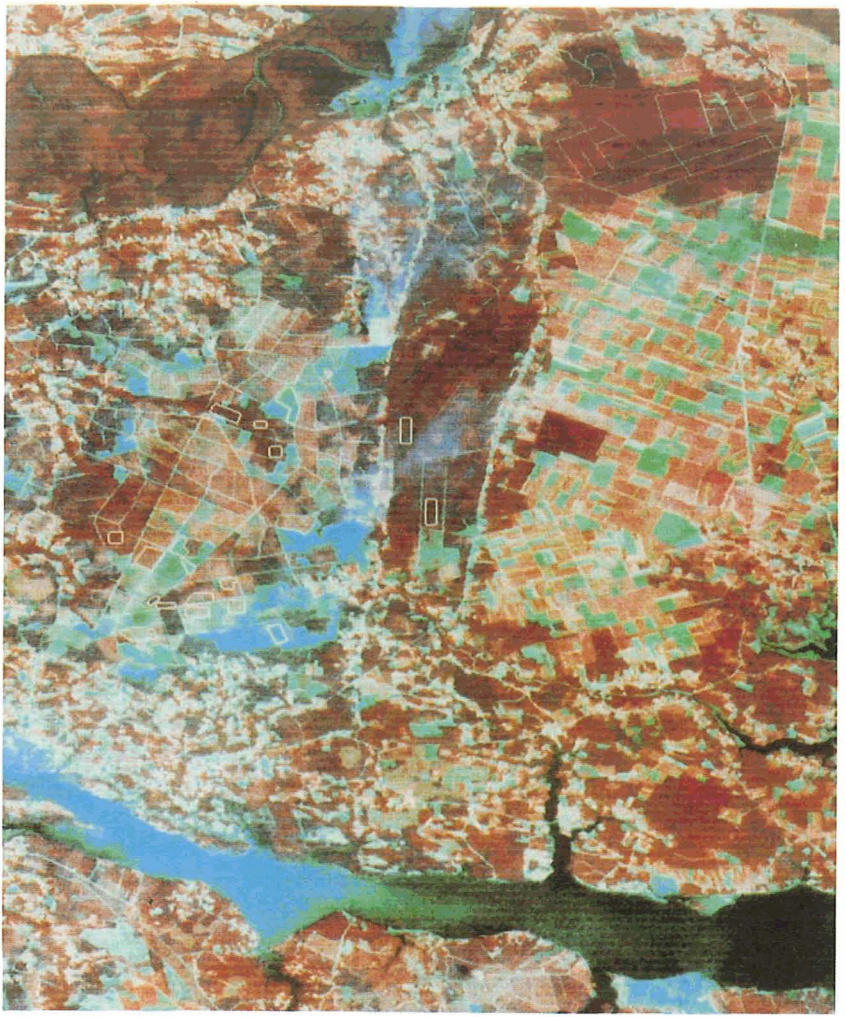


FIGURE 8.7—*Landsat MSS image with training areas for seven classes outlined (scene 1538–15100).*

MSS multispectral image of the same area taken on August 30, 1973, are shown in the right column of table 8.3. The maximum average divergences D_A for several feature subsets are shown in table 8.4.

In summary, the objective of feature selection is to find a small number of variables that have a large relevance for classification. If the dimensionality reduction is pushed too far, however, significant information and therefore discriminating capability is lost. If, on the other hand, the

TABLE 8.1—*Class Names and Number of Training Vectors M_k for Forest Type Classification*

Class number	Class name	M_k
1	Closed canopy	548
2	Open canopy	642
3	Partial close	199
4	Regeneration	702
5	Hardwood/pine	469
6	Clearcut	692
7	Old/clearcut	125

dimensionality of the feature space is too large, the available training patterns will be so sparsely distributed that the estimation of the probability densities becomes very inaccurate. Consequently, the dimensionality of the feature space should only be reduced to a certain N, where $N<P$, at which point the class probability density $p(\mathbf{z}|S_k)$ of the features will be used for classification.

8.3 Supervised Classification

In supervised classification a partition of the available training set into K subsets s_k is known. The training features are used to determine the class boundaries such that the classification of unknown features results in a minimum error rate. Two approaches are available. In the statistical approach, the class boundaries are given by the parameters of probability densities. In the geometric approach, the boundaries are represented by the coefficients of discriminant functions.

Supervised classification consists of the following steps:

1. Determination of the number of classes of interest
2. Selection of training set and determination of class boundaries
3. Feature selection
4. Classification of new patterns based on class characteristics and selected features

(See fig. 8.8.)

Important and practically difficult problems are the determination of the number of classes K and the selection of a representative training set. An interactive system facilitates this task. Images may be displayed on the screen of a display device. The determination of K and the selection and evaluation of s_k are performed interactively.

TABLE 8.2—*Statistical Characteristics of Training Data in Figure 8.6 for Seven Classes and Four Features*

Parameter	Covariance matrix computed with equation											
	(8.22)				(8.28)				(8.34)			
Total covariance matrix (elements)	1.38	1.22	1.08	0.54	4.13	6.79	2.01	0.46	5.90	8.75	3.55	1.20
	1.22	2.14	1.45	0.86	6.79	13.66	3.43	0.84	8.75	16.39	4.85	1.40
	1.08	1.45	3.91	1.83	2.01	3.43	15.02	9.49	3.55	4.85	18.33	11.56
	0.54	0.86	1.83	1.66	0.46	0.84	9.49	7.2	1.20	1.43	11.56	8.55
Eigenvalues	6.26	1.7	0.67	0.46	23.67	15.02	0.77	0.55	30.15	17.38	0.98	0.67
Eigenvectors	0.32	0.51	−0.29	0.74	0.24	0.39	−0.40	0.80	0.30	−0.38	0.77	0.42
	0.44	0.67	0.22	−0.56	0.45	0.76	0.29	−0.36	0.48	−0.72	−0.47	−0.14
	0.73	−0.47	−0.46	−0.17	0.73	−0.38	−0.49	−0.28	0.70	0.43	0.23	−0.52
	0.41	−0.26	0.81	0.33	0.45	−0.36	0.72	0.40	0.43	0.39	−0.38	0.73
Principal components	1	2	3	4	1	2	3	4	1	2	3	4
Percent variance	68.9	18.7	7.3	5.1	59.2	37.5	1.9	1.4	61.3	35.4	2.0	1.3
Cumulative percent variance	68.9	87.6	94.9	100.0	59.2	96.7	98.6	100.0	61.3	96.7	98.7	100.0

TABLE 8.3—*Pairwise Transformed Divergences (D_T) for Seven Classes of Forest Types Versus Class Numbers*

Four features			Eight features		
S_j	S_k	$D_T(S_j,S_k)$	S_j	S_k	$D_T(S_j,S_k)$
3	4	100.0	3	7	100.0
3	7	100.0	3	6	100.0
1	4	100.0	3	4	100.0
1	7	100.0	2	6	100.0
1	6	100.0	1	7	100.0
3	6	99.99	1	6	100.0
4	5	99.94	1	4	100.0
2	7	99.91	2	7	100.0
2	6	99.87	4	6	99.99
2	4	99.74	4	5	99.98
5	7	98.54	5	6	99.97
7	6	98.35	2	4	99.95
5	6	91.29	5	7	99.49
1	5	87.12	3	5	96.88
4	7	85.46	1	2	95.48
3	5	82.84	2	5	95.31
2	5	66.76	6	7	95.14
6	7	63.20	4	7	93.80
1	2	57.73	1	5	93.62
2	3	54.38	1	3	86.37
1	3	12.50	2	3	76.28

TABLE 8.4—*Average Divergences of Various Feature Subsets*

Number of original measurements P	Number of features N	Feature subset [1]	Average divergence D_A
4	4	1, 2, 3, 4	90.4
	3	2, 3, 4	84.5
	3	1, 2, 4	83.3
	3	1, 2, 3	83.0
	3	1, 3, 4	76.8
	2	2, 4	[2] 82.2
	2	3, 4	[3] 52.9
8	8	1, 2, 3, 4, 5, 6, 7, 8	96.8
	6	2, 3, 4, 6, 7, 8	[2] 96.2
	4	2, 4, 7, 8	[2] 93.1
	2	2, 8	[2] 87.6

[1] Features 1, 2, 3, and 4 are MSS bands 4, 5, 6, and 7, respectively, of February 1974. Features 5, 6, 7, and 8 are MSS bands 4, 5, 6, and 7, respectively, of August 1973.

[2] Best.

[3] Worst.

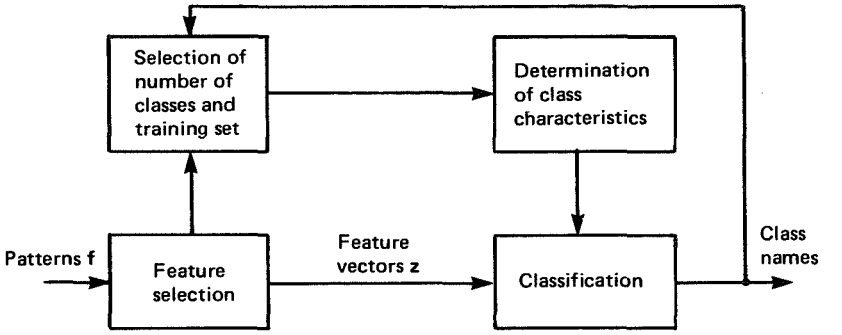


FIGURE 8.8—*Supervised classification.*

8.3.1 Statistical Classification

In the statistical approach the feature vectors $\mathbf{z}$ are considered random variables. Their description requires that the conditional probability density $p(\mathbf{z} \mid S_k)$ and the a priori probability $P(S_k)$ be known. Because of the homogeneity of objects, it may be assumed that the conditional probability of $\mathbf{z}$ depends only on the class to which the object containing $\mathbf{z}$ belongs. It is also assumed that the functional form of the distribution p is known and that only the parameters of p have to be determined from the training set. This process is referred to as parametric classification. Nonparametric techniques are used if the form of the underlying densities is unknown.

The design of a statistical classifier is based on a loss function, which evaluates correct and incorrect decisions. Let $p(\mathbf{z} \mid S_k)$ be the probability density for $\mathbf{z}$, given that $\mathbf{z}$ is from class S_k. Let $P(S_k)$ be the a priori probability of class S_k occurring. Let $\lambda(S_i \mid S_k)$ be the loss incurred when a pattern actually belonging to class S_k is assigned to class S_i. The conditional average loss $L(\mathbf{z}, S_k)$ is given by

$$L(\mathbf{z}, S_k) = \sum_{i \neq 1}^{K} \lambda(S_k \mid S_i) P(S_i \mid \mathbf{z}) \tag{8.42}$$

It is the loss associated with observing feature vector $\mathbf{z}$ and assigning it to various classes S_i, each weighted by the loss incurred by that particular classification. The a posteriori probability $P(S_i \mid \mathbf{z})$ is the probability of class S_i occurring having observed $\mathbf{z}$ [8, 9].

The classifier that minimizes the conditional average loss, given in equation (8.42), is called optimal or Bayesian classifier. Bayes' decision rule states that the classifier must assign $\mathbf{z}$ to the class S_k, where

$$L(\mathbf{z}, S_k) < L(\mathbf{z}, S_i) \text{ for all } i = 1, \ldots, K.$$

The a posteriori probability $P(S_i \mid \mathbf{z})$ can be computed from $p(\mathbf{z} \mid S_i)$ by the Bayes rule:

$$P(S_i \mid \mathbf{z}) = \frac{p(\mathbf{z} \mid S_i)P(S_i)}{\sum_{k=1}^{K} p(\mathbf{z} \mid S_k)P(S_k)} \tag{8.43}$$

Consider the following symmetric loss function:

$$\lambda(S_k \mid S_i) = \begin{cases} 0 & i=k \\ 1 & i \neq k \end{cases} \qquad i, k=1, \ldots, K \tag{8.44}$$

It assigns no loss to a correct decision and a unit loss to any error [9]. Thus, all errors are equally costly. The corresponding conditional average loss is $L(\mathbf{z}, S_k)$, where

$$L(\mathbf{z}, S_k) = \sum_{i=k} P(S_i \mid \mathbf{z}) = 1 - P(S_k \mid \mathbf{z}) \tag{8.45}$$

and $P(S_k \mid \mathbf{z})$ is the conditional probability that class S_k is correct. Thus, to minimize the conditional average loss, the class S_k that maximizes the a posteriori probability $P(S_k \mid \mathbf{z})$ should be selected.

A classifier can be represented by a set of discriminant functions $g_i(\mathbf{z})$, $i=1, \ldots, K$. Such a classifier assigns a feature vector $\mathbf{z}$ to class S_k if

$$g_k(\mathbf{z}) > g_i(\mathbf{z}) \qquad \text{for all } i \neq k \tag{8.46}$$

The classifier computes K discriminant functions and selects the class corresponding to the largest discriminant. A Bayesian classifier with symmetric loss function can now be represented by the discriminant functions:

$$g_k(\mathbf{z}) = P(S_k \mid \mathbf{z}) \qquad k=1, \ldots, K \tag{8.47}$$

It assigns a feature $\mathbf{z}$ to the class with the largest a posteriori probability. The choice of discriminant functions is not unique. If every $g_k(\mathbf{z})$ is replaced by $f(g_k(\mathbf{z}))$, where f is a monotonically increasing function, the resulting classification is unchanged. With use of equation (8.43), an equivalent representation of the Bayesian classifier is given by

$$g_k(\mathbf{z}) = p(\mathbf{z} \mid S_k)P(S_k) \tag{8.48}$$

Now the decision rule is: Given the feature $\mathbf{z}$, decide $\mathbf{z} \epsilon S_k$ if

$$p(\mathbf{z} \mid S_k)P(S_k) > p(\mathbf{z} \mid S_i)P(S_i) \qquad \text{for all } i \neq k \tag{8.49}$$

This criterion is commonly referred to as the maximum likelihood decision rule.

For remotely sensed images the assumption that $p(\mathbf{z} \mid S_k)$ is a multivariate normal (Gaussian) probability density distribution with mean $\mathbf{m}_k$ and covariance matrix $\mathbf{C}_k$ is often justified [11]. Thus:

$$p(\mathbf{z} \mid S_k) = \frac{1}{(2\pi)^{N/2} |\mathbf{C}_k|^{1/2}} e^{-1/2(\mathbf{z}-\mathbf{m}_k)^T \mathbf{C}_k^{-1}(\mathbf{z}-\mathbf{m}_k)} \tag{8.50}$$

where $|\mathbf{C}_k|$ is the determinant of $\mathbf{C}_k$. The mean vectors $\mathbf{m}_k$ are given by

$$\mathbf{m}_k = E\{\mathbf{z}_k\} \tag{8.51}$$

and covariance matrices $\mathbf{C}_k$ by

$$\mathbf{C}_k = E\{(\mathbf{z}-\mathbf{m}_k)(\mathbf{z}-\mathbf{m}_k)^T\} \tag{8.52}$$

are estimated from the M_k feature vectors ${}^j\mathbf{z}_k$ in each class of the training set:

$$\mathbf{m}_k \approx \frac{1}{M_k} \sum_{j=1}^{M_k} {}^j\mathbf{z}_k \tag{8.53}$$

and

$$\mathbf{C}_k \approx \frac{1}{M_k - 1} \sum_{j=1}^{M_k} ({}^j\mathbf{z}_k - \mathbf{m}_k)({}^j\mathbf{z}_k - \mathbf{m}_k)^T \tag{8.54}$$

The vectors ${}^j\mathbf{z}_k$ represent the training patterns, where k indexes the particular class and j indicates the jth prototype of class S_k. There may be M_k prototypes that are descriptive of the kth class S_k. Taking the logarithm of equation (8.48) and eliminating the constant term yield for a new g_k

$$g_k(\mathbf{z}) = \ln P(S_k) - \frac{1}{2} \ln |\mathbf{C}_k| - \frac{1}{2}(\mathbf{z}-\mathbf{m}_k)^T \mathbf{C}_k^{-1}(\mathbf{z}-\mathbf{m}_k) \tag{8.55}$$

Thus, for normally distributed patterns, the optimal classifier is a quadratic classifier.

Some of the assumptions made in the derivation of the maximum likelihood decision rule are often not realized in remotely sensed data:

1. The data from each class are normally distributed. This assumption has been shown to be erroneous at the 1-percent level of significance with a chi-square test on various data sets [11]. However, the assumption performs sufficiently well, and the use of a more complicated decision rule is not justified. Rather, radiometric errors and misregistration [12] should be corrected by preprocessing.
2. Class mean vectors and covariance matrices can be estimated from training data. The training set may not adequately describe the statistics of the classes if the number of measurements in s_k is insufficient, if a class is composed of subclasses, if the atmospheric conditions and the Sun and sensor positions relative to a ground

resolution element are different for training and nontraining data, and if the sensor generates noise (e.g., striping). Some of these errors can be removed by radiometric correction for haze, illumination, and sensor effects. (See sec. 3.2.)

3. The loss functions λ and the a priori probabilities $P(S_i)$ are known. These functions, however, cannot be accurately estimated.

Despite the radiometric corrections performed during preprocessing, there will be pattern vectors that do not belong to any of the classes defined by the training set. In remotely sensed images, such picture elements may represent roads, small water bodies, and mixtures of object points. The classification procedure assigns these patterns to one of the training classes, although they may yield very small discriminant values $g_k(\mathbf{z})$ for all classes. In the one-dimensional two-class example in figure 8.9, the patterns having a low probability of belonging to any of the training classes may be assigned to a reject class S_0 [7]. This operation can be performed by computing the probability density value associated with the feature vector and rejecting the point if the value is below a specified threshold.

Alternatively, the discriminant values stored as part of the classification result can be used. If $\mathbf{z}$ is N dimensional and normally distributed, the quadratic form $Q_k(\mathbf{z})$ has a chi-square (χ^2) distribution with N degrees of freedom ($c_N(\chi^2)$), where $Q_k(\mathbf{z})$ is given by

$$Q_k(\mathbf{z}) = (\mathbf{z}-\mathbf{m}_k)^T\mathbf{C}_k^{-1}(\mathbf{z}-\mathbf{m}_k) \tag{8.56}$$

Therefore thresholding r percent of the normal distribution P shown in figure 8.10 is equivalent to thresholding r percent of the chi-square distribution of $Q_k(\mathbf{z})$. The quadratic form $Q_k(\mathbf{z})$ is related to $g_k(\mathbf{z})$ in equation (8.55) in the following manner:

$$Q_k(\mathbf{z}) = -2g_k(\mathbf{z}) - \ln|\mathbf{C}_k| - 2\ln P(S_k) \tag{8.57}$$

Thus, every pattern with the following condition is assigned to the reject class S_0:

$$Q_k(\mathbf{z}) > \chi^2 \qquad \text{for which} \qquad c_N(\chi^2) = \frac{r}{100} \tag{8.58}$$

A different threshold value may be applied to each class.

Figure 8.11 shows the classification map obtained by classifying the image in figure 8.7 with the Bayesian classifier as given in equations (8.46) and (8.55). Feature vectors with low probability of correct classification are assigned to the reject class S_0, which is displayed in white. The large size of S_0 is due to the classifier being designed to recognize only forest areas, although the image also contains water and agricultural areas, which are assigned to S_0.

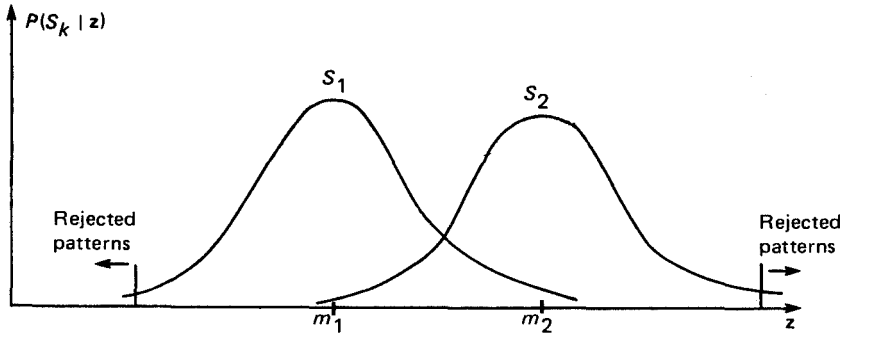


FIGURE 8.9—*Rejection of patterns with low probability of correct classification.*

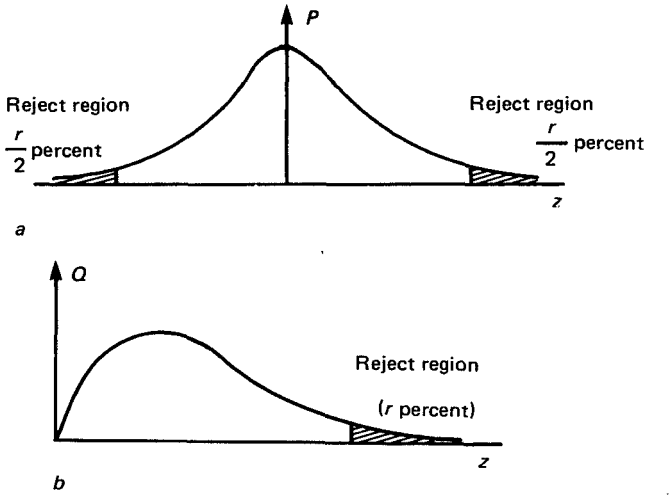


FIGURE 8.10—*Reject regions. (a) Normal distribution. (b) Chi-square distribution.*

The direct implementation of equation (8.46) with the discriminant function given in equation (8.55) results in a classifier that must compute this discriminant function for all classes for every pixel. The classification time is proportional to N^2K, where N and K are the dimension of the feature vector and the number of classes, respectively. However, the maximum likelihood classifier may be effectively implemented by a table lookup technique, because the number of unique feature vectors in remote sensing images is often approximately 10 percent of the total number of pixels [10]. Specifying a threshold value for each class defines the decision regions as hyperellipsoids. (See equation (8.10).) The boundary points

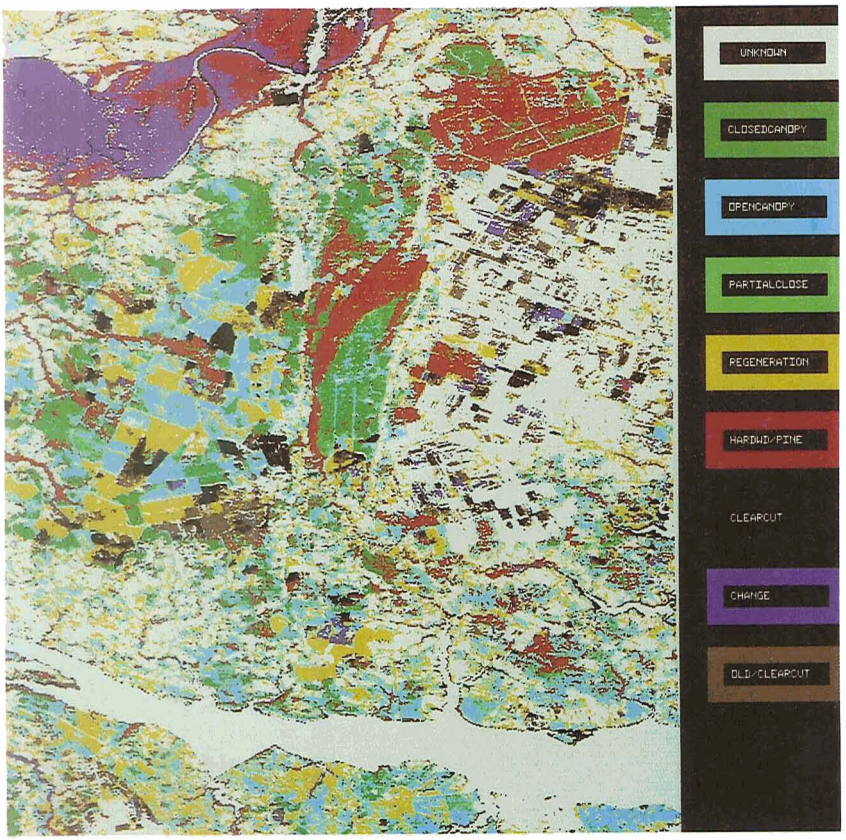


FIGURE 8.11—*Classification map obtained by maximum likelihood classification of image in figure 8.7 into seven classes.*

of these hyperellipsoids are computed and stored in lookup tables for later classification.

Building the tables and classifying an image by looking up the prestored class names for a feature vector requires considerably less computer time than classification with the direct method. In determining the boundary for a particular class, only a localized region of the feature space has to be searched, but in the direct implementation, all classes must be considered for every pixel. Figure 8.12 compares the classification times for both methods for 7 and 20 classes with 4 features used. The time is given in seconds of central processing unit (CPU) time measured for assembly language programs on an IBM 360/91 computer.

A further increase in computational speed may be obtained by approximation of the hyperellipsoid decision boundaries by hyperrectangles or parallelepipeds. Figure 8.13 shows the decision boundaries obtained from

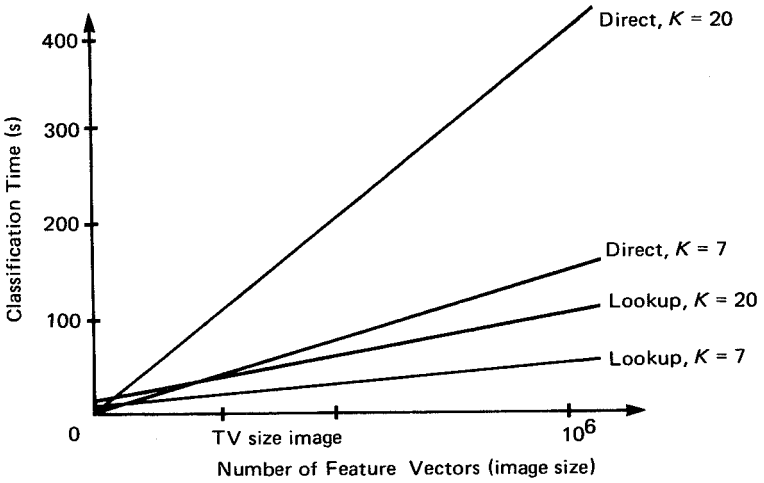


FIGURE 8.12—*Comparison of classification times for direct and table look-up implementation of maximum likelihood classifier for four features. (K is the number of classes.)*

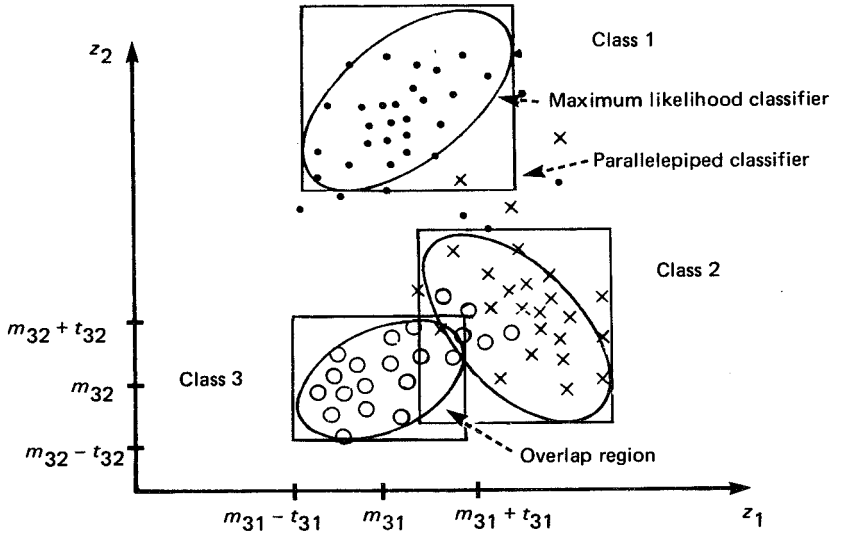


FIGURE 8.13—*Maximum likelihood and parallelepiped decision boundaries.*

equation (8.55) by holding $g_k(\mathbf{z})$ constant and their parallelepiped approximations for a three-class, two-feature problem. A feature vector $\mathbf{z}=(z_1, z_2, \ldots, z_N)^T$ is assigned to class S_k if

$$m_{ki}-t_{ki}\leq z_i\leq m_{ki}+t_{ki} \tag{8.59}$$

where m_{ki} and t_{ki} are the mean and threshold values of feature z_i for class S_k, respectively. If the parallelepipeds overlap, no unambiguous decisions are possible. Addington [13] proposed a hybrid classifier that uses the Bayesian decision rule to resolve ambiguities.

Currently most classifiers are implemented in computer programs. Hardware implementations are only known for simple decision rules such as the parallelepiped and the maximum likelihood classifiers. Because of the variability of remotely sensed data, a derived set of discriminant functions may not be used for different images. Classifiers usually have to be designed for each new image to be classified with training data.

8.3.2 Geometric Classification

Statistical classifiers are based on decision theory, and it can be shown that a given loss function is minimized. The drawback is that the conditional probability densities $p(\mathbf{z}|S_k)$ of the feature vectors must be known. For practical cases the determination of the multivariate probability densities is only an approximation to the real situation. Therefore, attempts are made to avoid random processes as models for the pattern classes.

In the geometric approach feature vectors $\mathbf{z}$ are viewed as points in the N-dimensional feature space. If the points $\mathbf{z}_k$ belonging to classes S_k, $k=1, \ldots, K$, form clusters, classification is achieved by finding decision surfaces that separate points of class S_k from those of classes S_j, $j=1, \ldots, K$, $j\neq k$. Construction of these surfaces is possible without knowledge of $p(\mathbf{z}|S_k)$. However, the determination of hypersurfaces may be as difficult as the determination of multivariate probability densities. A computationally feasible solution is obtained only if the separating decision surfaces are hyperplanes. The justification for this assumption depends essentially on the characteristics of the feature vectors.

The geometric approach is usually considered for the two-class problem. The feature space is divided by the hyperplane into two regions, one containing points of class S_1, the other containing all points not belonging to class S_1. The problem is to determine the equation of a hyperplane that optimally separates the classes S_1 and S_2, by using the classified feature vectors in the training set s, where

$$s=\{s_1, s_2\} \tag{8.60}$$

The decision surface is given by a discriminant function $g(\mathbf{z})$. The equation $g(\mathbf{z})=0$ defines the surface that separates points assigned to S_1 from points assigned to S_2. The problem of finding the discriminant functions can be formulated as a problem of minimizing a criterion function, which gives the average loss incurred in classifying the set of training vectors. Because it is very difficult to derive the minimum-risk linear discriminant, an iterative gradient descent procedure is used for minimizing the criterion function.

A linear discriminant function is given by

$$g(\mathbf{z})=w_0+\sum_{i=1}^{N} w_i z_i = w_0+\mathbf{w}^T\mathbf{z} \tag{8.61}$$

where w_i is the ith component of the weight factor, w_0 is the threshold weight, and N is the dimension of the feature vector $\mathbf{z}$. The quantity z_i is the ith component of the feature vector. A two-class linear classifier implements the following decision rule: If $g(\mathbf{z})>0$, decide class S_1; and if $g(\mathbf{z})<0$, decide class S_2.

Thus, $\mathbf{z}$ is assigned to class S_1 if the inner product $\mathbf{w}^T\mathbf{z}$ exceeds the threshold $-w_0$. The equation

$$g(\mathbf{z})=0 \tag{8.62}$$

defines the decision surface that separates points assigned to class S_1 from points assigned to class S_2. Because $g(\mathbf{z})$ is linear, this decision surface is a hyperplane H.

The discriminant function given in equation (8.61) gives an algebraic measure of the distance from $\mathbf{z}$ to the hyperplane. Vector $\mathbf{z}$ can be expressed as

$$\mathbf{z}=\mathbf{z}_p+r\frac{\mathbf{w}}{||\mathbf{w}||} \tag{8.63}$$

where $\mathbf{z}_p$ is the normal projection of $\mathbf{z}$ onto H and r is the distance between $\mathbf{z}$ and H. Then,

$$r=\frac{g(\mathbf{z})}{||\mathbf{w}||} \tag{8.64}$$

and the distance from the origin to H is given by $r_0=w_0/||\mathbf{w}||$. (See fig. 8.14.)

The linear discriminant function given in equation (8.61) can be written in homogeneous form

$$g(\mathbf{z})=\mathbf{a}^T\mathbf{y} \tag{8.65}$$

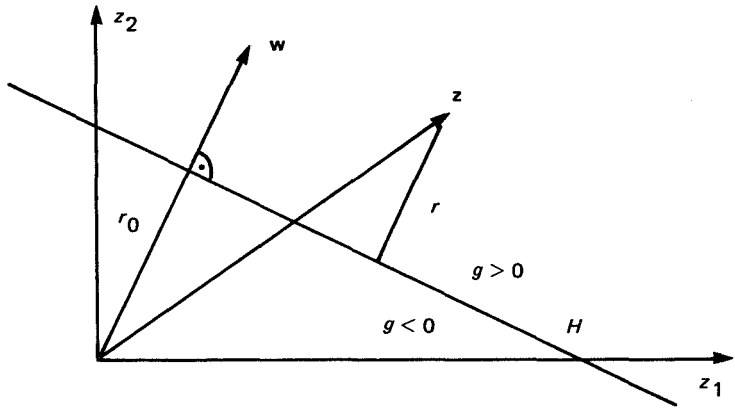


FIGURE 8.14—*Linear decision boundary.*

where

$$\mathbf{y}=\begin{pmatrix}1\\ z_1\\ \cdot\\ \cdot\\ \cdot\\ z_N\end{pmatrix} \quad \text{and} \quad \mathbf{a}=\begin{pmatrix}w_0\\ w_1\\ \cdot\\ \cdot\\ \cdot\\ w_N\end{pmatrix} \tag{8.66}$$

Given two distinct classes of patterns, the classifier design reduces to the problem of finding the weight vector **a**. Let ${}^j\mathbf{y}_1$, $j=1, \ldots, M_1$ be the feature vectors representing class S_1 (training vectors for class S_1), and let ${}^j\mathbf{y}_2$, $j=1, \ldots, M_2$ be the training vectors for class S_2. These vectors will be used to determine the weights in the linear discriminant function given in equation (8.65). If a solution exists for which all training vectors are correctly classified, the classes are said to be linearly separable.

A training vector ${}^j\mathbf{y}_k$ is classified correctly if $\mathbf{a}^T\,{}^j\mathbf{y}_1>0$ or if $\mathbf{a}^T\,{}^j\mathbf{y}_2<0$. Because $\mathbf{a}^T(-{}^j\mathbf{y}_2)>0$, replacing every ${}^j\mathbf{y}_2$ by its negative normalizes the design problem to finding a weight vector **a** such that

$$\mathbf{a}^T\,{}^j\mathbf{y}>0 \tag{8.67}$$

for all $M=M_1+M_2$ training vectors. Such a vector **a** is called a separating vector [8].

The weight vector **a** can be thought of as specifying a point in weight space. Each training vector ${}^j\mathbf{y}$ places a constraint on the possible location

of a separating vector. The equations $\mathbf{a}^{T}{}^{j}\mathbf{y}=0$ define hyperplanes through the origin of weight space, having ${}^{j}\mathbf{y}$ as normal vectors. The separating vector, if it exists, must be on the positive side of every hyperplane. Thus, it must be in the intersection of M half spaces, and any vector in this solution region is a separating vector. (See fig. 8.15.) It is evident that the separating vector is not unique.

The approach taken to find a solution to the set of linear inequalities $\mathbf{a}^{T}{}^{j}\mathbf{y}>0$ is to define a criterion function $J(\mathbf{a})$ that is minimized if $\mathbf{a}$ is a separating vector. This step reduces the problem to one of minimizing a scalar function that can be solved by a gradient descent procedure. It is important that the iterative procedure used does not converge to a limit point on the boundary. This problem can always be avoided by introducing a margin, i.e., by requiring

$$\mathbf{a}^{T}{}^{j}\mathbf{y}\geq\mathbf{b}>0 \qquad \text{for all } j=1,\ldots,M \tag{8.68}$$

The treatment of simultaneous linear equations is simplified by introducing matrix notation [9]. Let $\mathbf{Y}$ be the M by $(N+1)$ matrix whose jth row is the vector ${}^{j}\mathbf{y}^{T}$, and let $\mathbf{b}$ be the column vector $\mathbf{b}=(b_1,\ldots,b_M)^{T}$. Then the problem is to find a weight vector $\mathbf{a}$ satisfying

$$\mathbf{Ya}\geq\mathbf{b}>0 \tag{8.69}$$

The matrix $\mathbf{Y}$ is rectangular with more rows than columns. The vector $\mathbf{a}$ is overdetermined and can be computed by minimizing the error between $\mathbf{Ya}$ and $\mathbf{b}$.

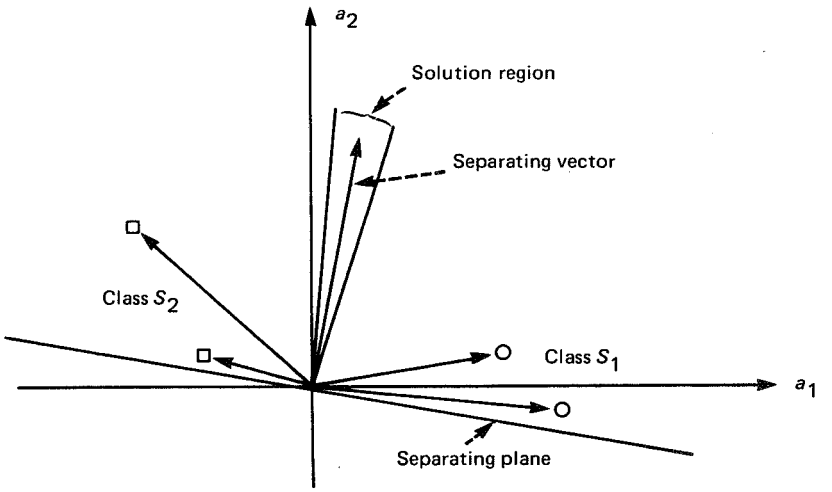


FIGURE 8.15—*Linearly separable training samples in weight space (from Duda and Hart [8]).*

The criterion function to be minimized is

$$J(\mathbf{a}, \mathbf{b}) = \tfrac{1}{2}||\mathbf{Ya}-\mathbf{b}|| \tag{8.70}$$

where $\mathbf{a}$ and $\mathbf{b}$ are allowed to vary subject to the constraint $\mathbf{b}>0$. The $\mathbf{a}$ that achieves the minimum is a separating vector if the training vectors are linearly separable.

To minimize J, a gradient descent procedure is used. The gradient of J with respect to $\mathbf{a}$ is given by

$$\nabla_{\mathbf{a}} J = \mathbf{Y}^T(\mathbf{Ya}-\mathbf{b}) \tag{8.71}$$

and the gradient of J with respect to $\mathbf{b}$ is given by

$$\nabla_{\mathbf{b}} J = -(\mathbf{Ya}-\mathbf{b}) \tag{8.72}$$

For a given $\mathbf{b}$

$$\mathbf{a} = (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{b} \tag{8.73}$$

thereby minimizing J with respect to $\mathbf{a}$. In modifying $\mathbf{b}$, the constraint $\mathbf{b}>0$ must be respected. The vector $\mathbf{b}$ is determined iteratively by starting with $\mathbf{b}>0$ and preventing $\mathbf{b}$ from converging to zero by setting all positive components of $\nabla_{\mathbf{b}} J$ to zero. This is the Ho–Kashyap algorithm [14] for minimizing $J(\mathbf{a}, \mathbf{b})$, summarized as follows:

$$\left.\begin{aligned} \mathbf{b}(0) &> 0, \text{ otherwise arbitrary} \\ \mathbf{a}(0) &= (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{b}(0) \\ \mathbf{b}(i+1) &= \mathbf{b}(i)+\rho[\mathbf{e}(i)+|\mathbf{e}(i)|] \\ \mathbf{a}(i+1) &= (\mathbf{Y}^T\mathbf{Y})^{-1}\mathbf{Y}^T\mathbf{b}(i+1) \\ i &= 0, 1, \ldots \end{aligned}\right\} \tag{8.74}$$

with the error vector

$$\mathbf{e}(i) = \mathbf{Ya}(i)-\mathbf{b}(i) \tag{8.75}$$

If the training samples are linearly separable, the Ho–Kashyap algorithm yields a solution in a finite number of steps with $\mathbf{e}$ converging to zero. If the samples are not linearly separable, $\mathbf{e}$ converges to a nonzero value.

If K classes are present, linear classifiers may be designed as parallel or sequential classifiers. A parallel classifier requires $K(K-1)$ segments of linear discriminant functions, one for every pair of classes. A sequential classifier is structurally simpler, because the K-class problem is reduced to $K-1$ two-class problems that are solved sequentially. Thus, a sequential classifier makes a series of pairwise decisions. At the kth stage a linear discriminant function separates feature vectors assigned to class S_k from those not assigned to class S_k [8, 15]. The classification time is proportional to $N(K-1)$. Because an iterative technique is used, the time to design the linear sequential classifier is dependent on the size and characteristics of the training set.

8.4 Unsupervised Classification

Supervised classification techniques require a set of training patterns whose class membership is known. If a labeled or classified training set is not available, unsupervised classification techniques must be used. The only available knowledge is that the patterns in the given training set s belong to one of K classes. Additional problems are encountered if the number of classes K is also unknown.

There are several reasons why unsupervised classification is of great practical importance. The determination of the number of object classes in an image and the determination of a labeled training set for supervised classification often present practical difficulties. In the early stages of image analysis, unsupervised classification is valuable to gain insight into the structure of the data. Image patterns often change slowly with time. These temporal changes may be tracked by an unsupervised classifier. Two approaches to unsupervised classification may be distinguished: statistical and clustering. Unsupervised classification is only possible if assumptions about the structure of the patterns can be made. This structure is reflected in conditional probability density distributions or in similarity measures. Classification is impossible if the feature vectors are randomly distributed in feature space.

8.4.1 Statistical Unsupervised Classification

In the statistical approach, the probability density function for the feature vectors $\mathbf{z}$ in the unlabeled training set s is the mixture density $p(\mathbf{z})$, given by

$$p(\mathbf{z}) = \sum_{k=1}^{K} P(S_k) p(\mathbf{z} \mid S_k) \tag{8.76}$$

Some or all of the quantities $\{K, P(S_k), p(\mathbf{z} \mid S_k), k=1, \ldots, K\}$ may be unknown. Thus, unsupervised classification is the estimation of the unknown quantities in equation (8.76) with the feature vectors in s. No general solution of this problem is known. A solution exists under the assumption that $K, P(S_k)$ and the form of $p(\mathbf{z} \mid S_k)$ are known, and only the parameters of $p(\mathbf{z} \mid S_k)$ have to be estimated [8].

8.4.2 Clustering

Clustering does not assume any knowledge of probability density distributions of the feature vectors and formulates the problem as one of partitioning the patterns into subgroups or clusters. This approach is based on a measure of similarity. Thus, clustering is a technique for pattern classification in terms of groups of patterns or clusters that possess strong internal similarities. Viewed geometrically, the patterns form clouds of points in N-dimensional feature space. Clustering consists of two prob-

lems: (1) the definition of a measure of similarity between the patterns and (2) the evaluation of a partition of a set of patterns into clusters.

The two basic data characteristics that can be used as measures of similarity are distance between patterns in feature or pattern space and density of patterns (i.e., the number of points) in different regions of these spaces. The distance between patterns in the same cluster can be expected to be significantly less than the distance between patterns in different clusters. Alternatively, relatively dense regions in feature space, each of which is separated from the other by sparsely populated regions, can be regarded as the set of pattern clusters. More than one of the clusters may belong to the same pattern class, or patterns from different classes may contribute to one cluster.

The problem of clustering is to partition a set of M patterns or feature vectors $\mathbf{z}_1, \ldots, \mathbf{z}_M$ into K disjoint subsets s_i, where $\mathbf{z}_1, \ldots, \mathbf{z}_M$ compose the entire set s:

$$s=\{\mathbf{z}_1, \ldots, \mathbf{z}_M\} \tag{8.77}$$

Each subset is to represent a cluster, with patterns in one cluster being more similar than patterns in different clusters according to some similarity measure.

The Euclidean distance between two feature vectors $\mathbf{z}_i$ and $\mathbf{z}_j$

$$d=||\mathbf{z}_i-\mathbf{z}_j||=[(\mathbf{z}_i-\mathbf{z}_j)^T(\mathbf{z}_i-\mathbf{z}_j)]^{1/2} \tag{8.78}$$

may be used as a measure of similarity. However, clusters defined by the Euclidean distance will only be invariant to translations and rotations but not to linear transformations in general. Simple scaling of the coordinate axes, for example, can result in a different grouping of the patterns into clusters [8, 16].

Instead of a distance measure, the following nonmetric similarity function may be used:

$$S(\mathbf{z}_i, \mathbf{z}_j)=\frac{\mathbf{z}_i^T\mathbf{z}_j}{||\mathbf{z}_i||\,||\mathbf{z}_j||} \tag{8.79}$$

which is the cosine of the angle between the vectors $\mathbf{z}_i$ and $\mathbf{z}_j$. Use of this measure is governed by certain assumptions, such as sufficient separation of clusters with respect to the coordinate system origin.

After adoption of a similarity measure, a criterion function has to be defined that measures the clustering quality of any partition of the patterns. One possibility is to define a performance index and select the partition that extremizes this index. A simple criterion function for clustering is the sum of squared errors index

$$J=\sum_{k=1}^{K}\sum_{\mathbf{z}\in s_k}||\mathbf{z}-\mathbf{m}_k|| \tag{8.80}$$

where

$$\mathbf{m}_k = \frac{1}{M_k} \sum_{\mathbf{z} \in s_k} \mathbf{z} \tag{8.81}$$

i.e., $\mathbf{m}_k$ is the mean vector of cluster s_k. The quantity M_k is the number of patterns in s_k. Thus, for a given cluster s_k, the mean vector $\mathbf{m}_k$ is the best representative of the patterns in s_k in the sense that it minimizes the sum of the squared errors between the patterns of the cluster and its mean. An optimal partitioning is defined as one that minimizes the criterion function J. A clustering algorithm usually chooses an initial partition followed by an iterative procedure that reassigns feature vectors to clusters until an extremum of J is reached. Such a procedure can only guarantee a local extremum. Different starting points can lead to different solutions. This procedure is represented by the following basic clustering algorithm:

1. Select number of clusters K.
2. Choose an initial partition of the M feature vectors $\mathbf{z}$ into K clusters. Compute cluster means $\mathbf{m}_1, \ldots, \mathbf{m}_K$.
3. Calculate the distance of a feature vector $\mathbf{z}$ to each cluster mean.
4. Assign $\mathbf{z}$ to the nearest cluster.
5. If no reassignment occurs (algorithm converged), or if the maximum number of iterations is reached, stop.
6. Otherwise calculate new cluster means and go to 3.

The disadvantage of this clustering procedure is that the number of clusters must be known and that the number of distances to be computed increases with the square of the number of pattern vectors to be analyzed. Additionally, the assumption is made that the classes are nonoverlapping in feature space. Overlapping pattern classes can lead to clusters that are mixtures of different pattern classes.

An alternative to this concept of intersample distance measures for cluster development is the use of the density of samples. The sample density is used for parametric classification, where the distributions of the different pattern classes are known and their parameters are determined from a given set of training vectors. In a nonparametric unsupervised environment the regions of the feature space with higher density are regarded as the pattern clusters.

A clustering algorithm that learns the number of clusters from the data distribution and permits overlapping of classes as long as the sample density in the region of overlap is less than the sample densities in the neighboring nonoverlapping regions was described by Dasarathy [17, 18]. The sample densities are estimated from the multidimensional histogram of the multi-image. The clusters are developed from the multidimensional histogram by merging each cell in the histogram space with its higher

density neighbor. The algorithm identifies the hills and valleys in the histogram where some of the hills may be the result of overlapping distributions. Defining the centroids of such hills as cluster centers will lead to clusters that may represent a mixture of more than one pattern class. Not all cells determined by the merging process represent significant clusters.

Mean cluster density and intercluster distance are used to derive a measure of significance for each cluster. First, all candidate cells whose densities exceed the average density D_a before clustering are selected as definitely significant clusters. The average density is given by

$$D_a = \frac{M}{H} \tag{8.82}$$

where H is the number of nonempty cells in the original histogram, and M is the number of feature vectors to be analyzed. The distances between these clusters are then computed by

$$d_{ki} = || \mathbf{m}_k - \mathbf{m}_i || \tag{8.83}$$

where $\mathbf{m}_k$ is the centroid of cluster s_k and $\mathbf{m}_i$ is the centroid of the definitely significant cluster s_i closest to s_k. The quantity $d_{\max}$ is the maximum distance between the definitely significant clusters.

A measure of significance of a cluster s_k is defined as

$$q_k = \frac{d_{ki} M_k}{d_{\max} M_m} \tag{8.84}$$

where d_{ki} is the distance between cluster s_k and its nearest definitely significant cluster. The variable M_k is the population of cluster s_k, and M_m is the population of the highest density cluster. The minimum value of q_k over the set of all definitely significant clusters is defined as the acceptable level of significance

$$q = \min(q_k) \tag{8.85}$$

Any cluster s_j for which $q_j > q$ is accepted as significant. If $q_i < q$, cluster s_j is merged with its nearest definitely significant neighbor. Classification of the unknown patterns may be achieved with any classification algorithm with the information derived in the clustering process. A similar clustering algorithm is described in [19].

8.5 Classifier Evaluation

The performance of a classifier is evaluated in terms of its error rate. The calculation of the error rate is too difficult for multiclass problems. Therefore, classifiers are tested experimentally by using the fraction of the pattern vectors of a test set that are misclassified as an estimate of the

error rate. If the true error rate of a classifier is p_e, and if m of the M test patterns are misclassified, then m has a binomial distribution:

$$P(m)=\binom{M}{m}p_e^m(1-p_e)^{M-m} \tag{8.86}$$

Confidence intervals for this distribution are tabulated with the size of the test set M as parameter [8]. Unless M is fairly large, the estimate of probability $\hat{p}_e$ of correct classification, where

$$\hat{p}_e=\frac{m}{M} \tag{8.87}$$

must be interpreted with caution. For example, if the error estimate is 0.15, obtained with 100 test patterns, the true error rate is between 8 and 24 percent with probability 0.95.

The sizes of the training and test sets influence the estimated classifier performance. The training set s is used to estimate distribution parameters or parameters of a discriminant function. Theoretically this estimation requires an infinite size of s. On the other hand, the cost of selecting and processing a training and test set increases with its size. It has also been observed that with a finite size of s, classifier performance does often not improve when the number of features is increased. The relationship between a finite training set size and the dimensionality of the feature vectors has been investigated by Kanal [20].

The results of the comparison of two feature selection methods and their effect on classifier performance are summarized in table 8.5. The evaluation is based on the forest type classification problem with seven classes, introduced in section 8.2. Measurements from one Landsat MSS image with four spectral bands (shown in fig. 8.7) from February 26, 1974, and from a second MSS image taken on August 30, 1973, are used as patterns. The K–L expansion based on the three covariance matrices in equations (8.22), (8.28), and (8.34), and average transformed divergence are used for feature selection. The selected feature subsets are classified with a Bayesian classifier and compared with the performance of a linear classifier. Classifier performance is expressed as probability of correct classification and classification time. The time is given in seconds CPU time per 10^6 feature vectors on an IBM 360/91 computer.

To select the best N features from P original measurements with use of transformed divergence, $m=P!/[N!(P-N)!]$ possible feature combinations have to be evaluated. For the K–L transform the first N components are the best selection. Table 8.5 also compares the performance of the criteria used to derive the orthogonal transform. The covariance matrix in equation (8.22) describes the optimal representation of the pattern classes, but use of equation (8.34) maximizes the distance between feature vectors from different classes. For the test data, both criteria yield approximately the same performance and are inferior to

TABLE 8.5—*Feature Selection and Classifier Performance*

Classification method	Number of measurements	Number of features	Feature selection method	Correct classification (percent)	Classification time (seconds) 10^6 vectors)
Quadratic (maximum likelihood) classifier	4 (MSS image from February 26, 1974)	4	Original 4 MSS bands	78.8	160
		3	Divergence; MSS 5, 6, 7	77.8	156
		3	K–L components 1, 2, 3; eq. (8.22)	74.1	235
		3	K–L components 1, 2, 3; eq. (8.28)	66.6	235
		3	K–L components 1, 2, 3; eq. (8.34)	71.1	235
		2	Divergence; MSS 5, 7	76.0	115
		2	K–L components 1, 2; eq. (8.22) σ_t = 87.6 percent	74.1	176
	8 (Two MSS images from February 26, 1974 and August 30, 1973)	8	Original 2 × 4 MSS bands	92.1	353
		6	Divergence (MSS 5, 6, 7)[1] and (MSS 5, 6, 7)[2]	92.2	276
		6	K–L components 1, 2, 3, 4, 5, 6; eq. (8.22); σ_t = 96.2 percent	90.6	582
		4	Divergence (MSS 5, 7)[1] and (MSS 6, 7)[2]	90.5	178
		4	K–L components 1, 2, 3, 4; eq. (8.22); σ_t = 90.9 percent	89.1	376
		4	K–L components 1, 2, 3, 4; eq. (8.28); σ_t = 97.6 percent	81.1	376
		4	K–L component 1, 2, 3, 4; eq. (8.34); σ_t = 97.8 percent	86.1	376
		2	Divergence (MSS 5)[1] and (MSS 7)[2]	83.0	118
		2	K–L components 1, 2; eq. (8.22); σ_t = 68.1 percent	76.8	218
		2	K–L components 1, 2; eq. (8.28); σ_t = 87.2 percent	73.7	218
		2	K–L components 1, 2; eq. (8.34)	76.9	218
Linear classifier	4 (MSS image from February 26, 1974)	4	Original 4 MSS bands	71.5	157

Note: σ_t is percent variance represented by components.
[1] Data from February 26, 1974.
[2] Data from August 30, 1973.

feature selection with the average transformed divergence method in both correct classification and time. The time required to perform the K–L transform is considerably longer than the time to calculate the average divergences. The performance of features obtained with covariance matrix in equation (8.28) is significantly inferior. Figure 8.16 compares the probability of correct classification and classification time for K–L transform and transformed divergence. The graphical presentation uses values from table 8.5 for $P=8$ measurements and the covariance matrix in equation (8.22). The classification accuracies of the individual classes are listed in table 8.6. For the multitemporal data the best four channels are selected according to the transformed divergence criterion. The classification accuracies for classes 1, 2, and 3 using all four channels from one multispectral Landsat image are very low. The poor separability of these classes is reflected in the corresponding low divergences in table 8.3. A graphical representation of the estimate of the classification accuracies using four features is shown in figure 8.17. Selection of the best four features from the eight measurements of a multitemporal image yields an almost uniform accuracy for all classes at only a slight increase in classification time.

Most existing classifiers use only spectral and temporal information by classifying individual picture elements as pattern vectors. Classification accuracy can be increased by using spatial information. The combination of segmentation into regions and classification of each region as a unit rather than classifying each individual picture element allows the use of texture and other spatial characteristics of objects [21, 22].

Another problem affecting classifier performance in remote sensing applications is estimating the expected proportions of objects that cannot be observed directly or distinctly. Because of the limits in the spatial resolution of the instruments, different classes of objects may contribute to a single resolution element. With the radiation measured being a mixture of object classes, the pattern vectors are not characteristic of any object class [23, 24].

8.6 Classification Examples

Multispectral classification has been successfully applied as an image analysis technique for remote sensing studies in land use and agriculture [25]. In the field of geology, classification has not been as successful, primarily because of the nonhomogeneity of geologic units, presence of gradational boundaries, topographic effects, confusing influence of vegetative cover, and similarity of the spectral signature of different lithologies.

The results of classifying the Landsat MSS image in figure 4.15 with four spectral measurements as features are shown in figure 8.18. The thematic map obtained by supervised classification with the maximum likelihood classifier given by equation (8.55) into 20 training classes is

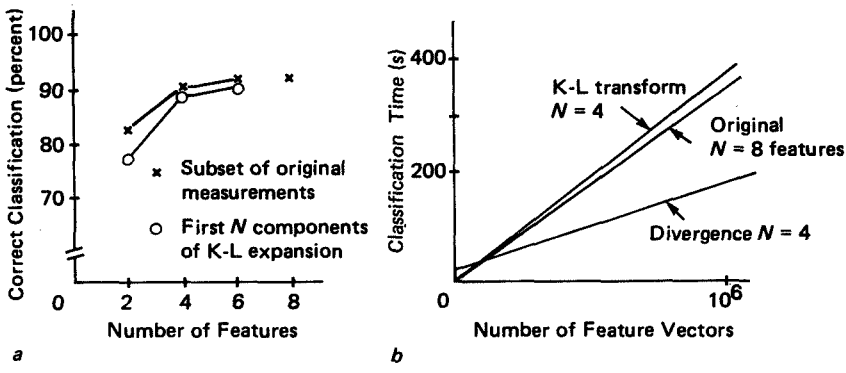


FIGURE 8.16—*Effect of feature selection on classification accuracy and classification time. (a) Classification accuracy versus number of features with original spectral bands and principal components. (b) Classification time versus number of feature vectors classified (360/91 CPU time).*

TABLE 8.6—*Correct Classification of Individual Classes*

Class	Correct classification (percent)		
	Single image	Multitemporal image	
	$P = 4, N = 4$	$P = 8, N = 4$	$P = 8, N = 8$
1	68.6	87.8	90.1
2	76.7	90.2	92.1
3	54.3	90.5	88.5
4	93.7	92.7	95.5
5	87.0	89.2	90.6
6	88.5	95.7	96.7
7	82.4	87.2	91.2
Total	78.8	90.5	92.1

shown in figure 8.18*a*. Figure 8.18*b* is the result of applying the clustering technique described in section 8.4.2 to the same image. Twenty significant clusters were obtained by analysis of the four-dimensional histogram. Geologists agree that the color enhancement of ratio images shown in figure 4.15*c* is superior to the classification results.

An example of supervised classification as a technique for the analysis of remotely sensed images using bipolarized rather than multispectral measurements is the detection of rainfall areas over land. Remote sensing of precipitation is fundamental to weather, climate, and Earth resources research activities. Upwelling microwave radiation measured by the Nimbus–6 Electrically Scanning Microwave Radiometer (ESMR–6) can

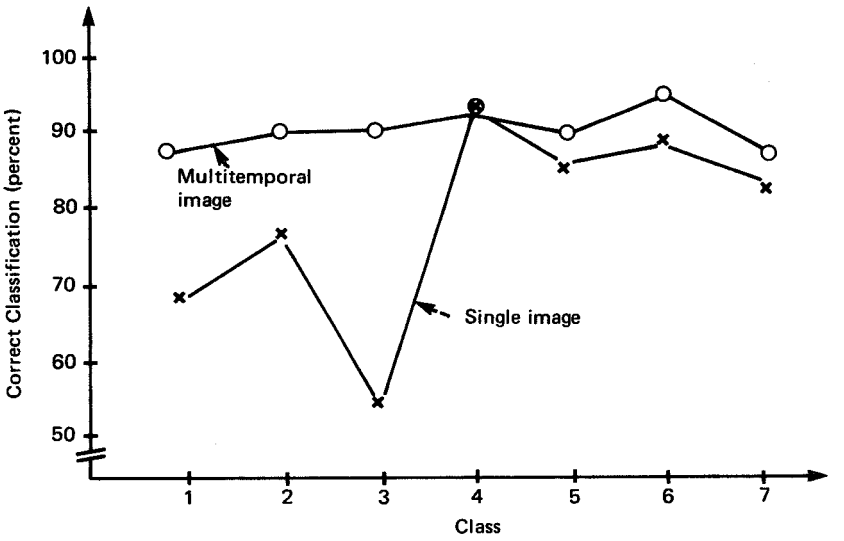


FIGURE 8.17—*Class classification accuracies using four features.*

be used to distinguish areas with rain over land and over ocean from areas with dry ground, with moist soil, or with no rain over ocean [26]. The ESMR–6 system measures thermal microwave radiation upwelling from the Earth's surface and atmosphere in a 250-MHz band centered at 37 GHz in two orthogonal (horizontal and vertical) polarizations [27]. The spatial resolution is approximately 20 by 40 km.

The problem is to classify two-dimensional feature vectors (horizontal and vertical brightness temperatures) into five classes. The training set required to design the classifier was derived from the ESMR–6 data by using radar and ground station measurements coinciding with the Nimbus–6 overpass. Figure 8.19 shows a scatter plot of the horizontal and vertical polarized brightness temperatures for the five classes. With the assumption of normally distributed data, the ellipses represent the decision boundaries defined by equation (8.55) for $r=32$ (i.e., 68 percent of the data within a class population—the data within one standard deviation—are encompassed by each ellipse). Pattern vectors outside the ellipses are assigned to the reject class S_0.

The lines represent the linear decision boundaries obtained for the linear classifier given by equation (8.60) with sequential decisions. No rain over ocean areas is first separated from rain over ocean, dry ground, wet soil, and rain over land areas. Next, rain over ocean areas is separated from the remaining classes. Then dry ground areas are separated from the two classes most difficult to separate: wet ground and rain over land areas. A large overlap occurs between data obtained from rainfall

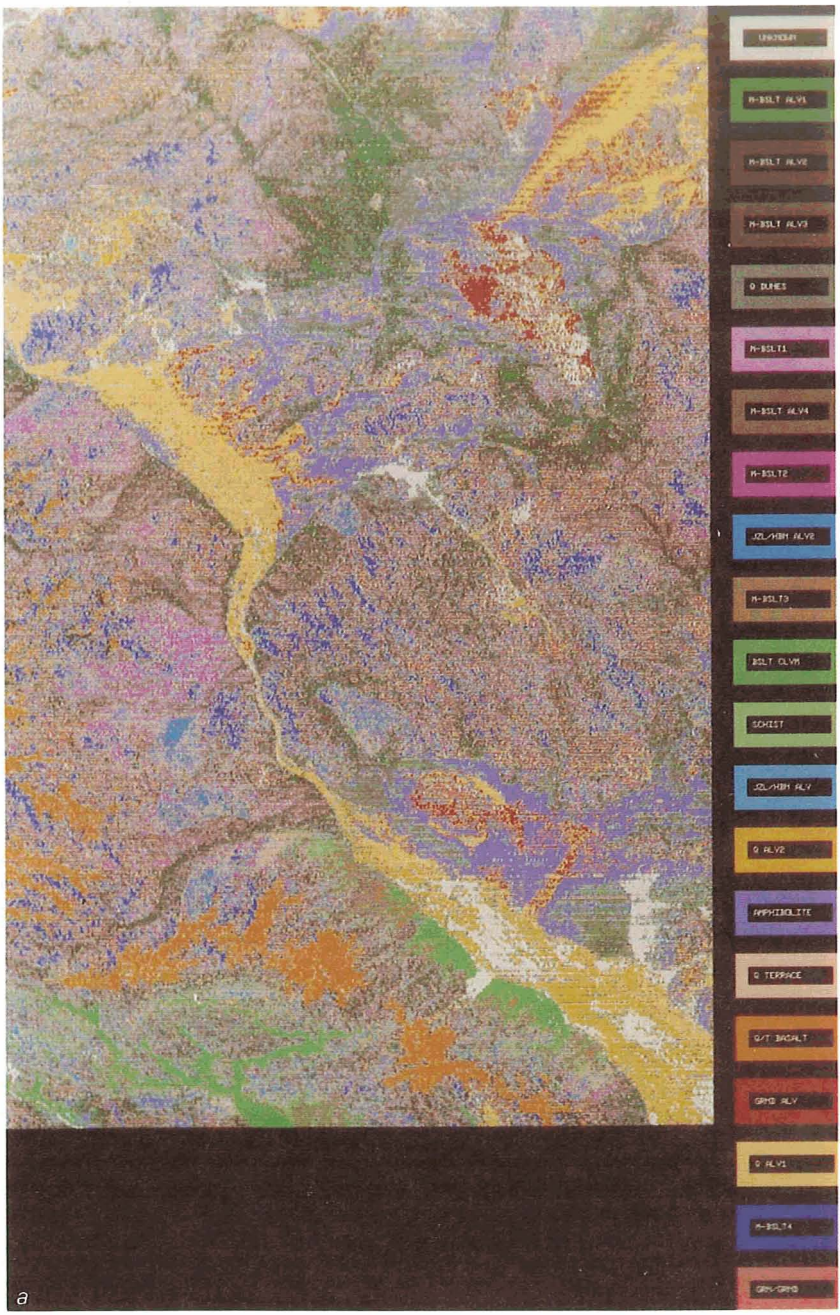


FIGURE 8.18a—*Map obtained by supervised classification into 20 classes.*

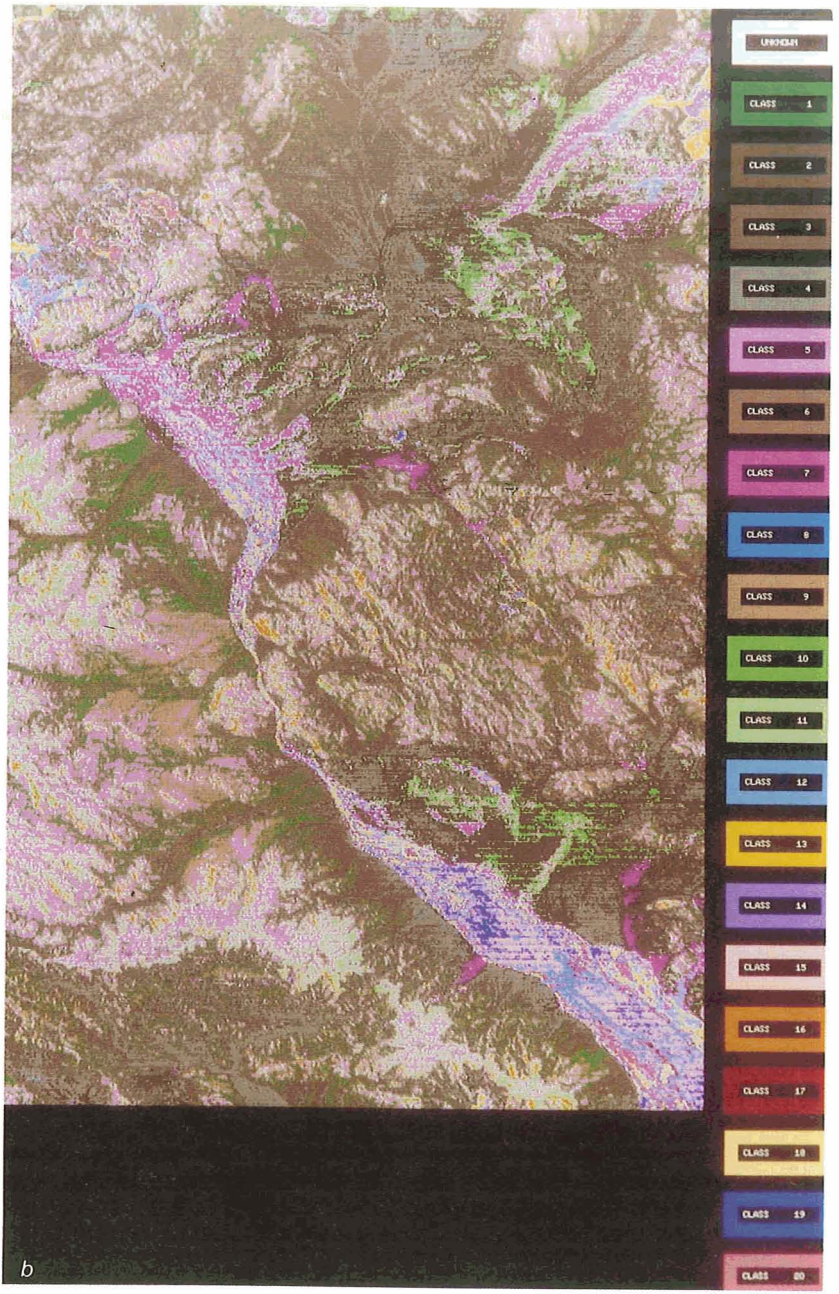


FIGURE 8.18*b*—*Map obtained by clustering.*

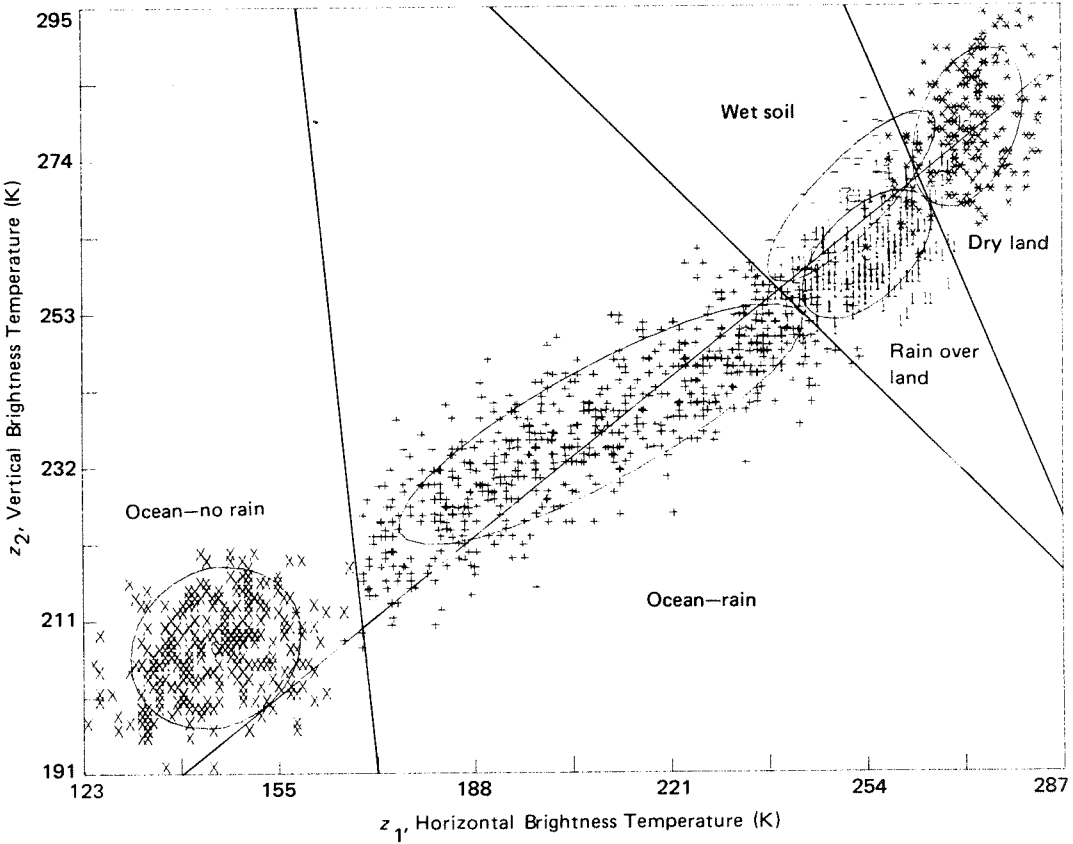


FIGURE 8.19—*ESMR-6 horizontally polarized versus vertically polarized brightness temperatures.*

over land areas and wet ground surfaces. Consequently, these two classes are difficult to separate. The results of a chi-square test [28] show that the assumption of a normal distribution of the class populations is justified. The classification map obtained with a Bayesian classifier for an area over the Southeastern United States is shown in figure 8.20. The geometric distortions caused by the conical scanner were not corrected.

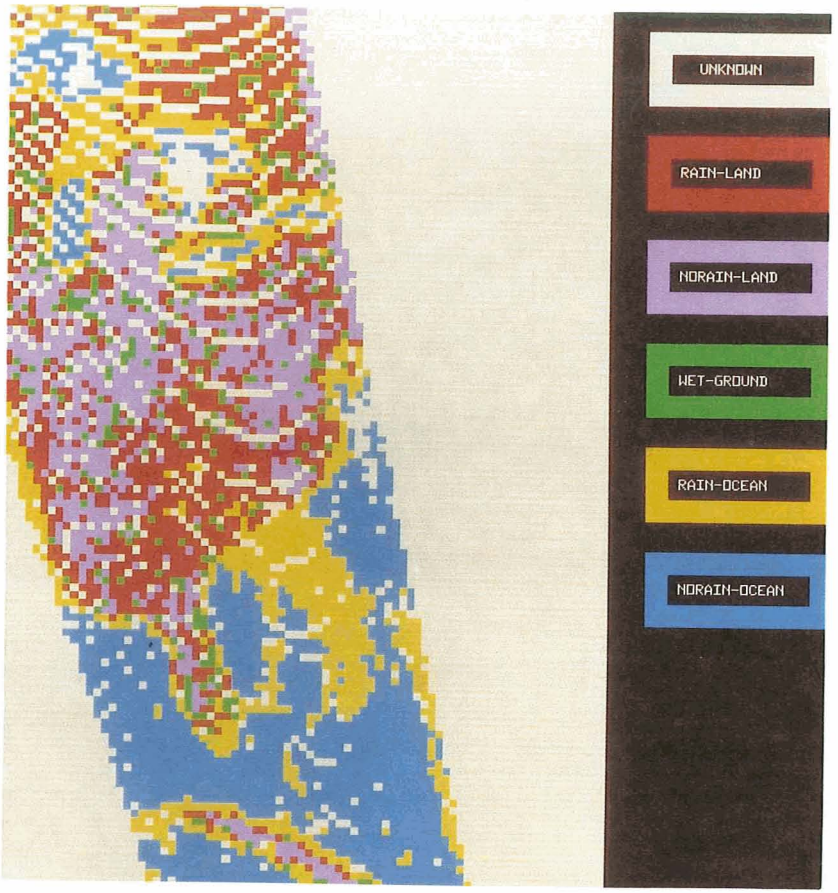


FIGURE 8.20—*Rain classification map from ESMR-6 data.*

REFERENCES

[1] Patrick, E. A.: Interactive Pattern Analysis and Classification Utilizing Prior Knowledge, Pattern Recognition, vol. 3, 1971, pp. 53–71.
[2] Tou, J. T.; and Heydorn, R. P.: Some Approaches to Optimum Feature Extraction, *in* Tou, J., ed.: Computers and Information Sciences—II. Academic Press, New York, 1967.
[3] Watanabe, S., et al.: Evaluation and Selection of Variables in Pattern Recognition, *in* Tou, J., ed.: Computers and Information Sciences—II. Academic Press, New York, 1967.
[4] Kailath, T.: The Divergence and Battacharyya Distance Measures in Signal Detection, IEEE Trans. Commun. Technol., vol. 15, no. 1, 1967, pp. 52–60.
[5] Chien, Y. T.; and Fu, K. S.: On the Generalized Karhunen-Loève Expansion, IEEE Trans. Inf. Theory, vol. IT-13, 1967, pp. 518–520.
[6] Swain, P. H.; and King, R. C.: Two Effective Feature Selection Criteria for Multispectral Remote Sensing. LARS Information Note 042673, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, Ind., 1973.
[7] Swain, P. H.: Pattern Recognition: A Basis for Remote Sensing Data Analysis. LARS Information Note 111572, Laboratory for Applications of Remote Sensing, Purdue University, Lafayette, Ind., 1973.
[8] Duda, R. D.; and Hart, P. E.: Pattern Classification and Scene Analysis. Wiley-Interscience, New York, 1973.
[9] Andrews, H. C.: Mathematical Techniques in Pattern Recognition. Wiley-Interscience, New York, 1972.
[10] Eppler, W. G.: An Improved Version of the Table Look-Up Algorithm for Pattern Recognition. Ninth International Symposium on Remote Sensing of the Environment, Ann Arbor, Mich., 1974, pp. 793–812.
[11] Crane, R. B.; Malila, W. A.; and Richardson, W.: Suitability of the Normal Density Assumption for Processing Multispectral Scanner Data, IEEE Trans. Geosci. Electron., vol. GE-10, 1972, pp. 158–165.
[12] Cicone, R. C.; Malila, W. A.; Gleason, J. M.; and Nalepka, R. F.: Effects of Misregistration on Multispectral Recognition. Proceedings of Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Lafayette, Ind., 1976, pp. 4A-1–4A-8.
[13] Addington, J. D.: A Hybrid Classifier Using the Parallelepiped and Bayesian Techniques. Proceedings of the American Society of Photogrammetry, Mar. 1975, Washington, D.C., pp. 772–784.
[14] Ho, Y. C.; and Kashyap, R. L.: A Class of Iterative Procedures for Linear Inequalities, SIAM J. Control, vol. 4, 1966, pp. 112–115.
[15] Bond, A. D.; and Atkinson, R. J.: An Integrated Feature Selection and Supervised Learning Scheme for Fast Computer Classification of Multispectral Data. Conference on Earth Resources Observation and Information Analysis Systems, University of Tennessee, Knoxville, Tenn., Mar. 1972.
[16] Tou, J. T.; and Gonsalez, R. C.: Pattern Recognition Principles. Addison-Wesley, Reading, Mass., 1974.
[17] Dasarathy, B. V.: An Innovative Clustering Technique for Unsupervised Learning in the Context of Remotely Sensed Earth Resources Data Analysis, Int. J. Syst. Sci., vol. 6, 1975, pp. 23–32.
[18] Dasarathy, B. V.: HINDU—Histogram Inspired Neighborhood Discerning Unsupervised System of Pattern Recognition: System Concepts. Computer Sciences Corp. Memo., 5E3080-4-8, July 1976.
[19] Goldberg, M.; and Shlien, S.: A Clustering Scheme for Multispectral Images, IEEE Trans. Systems, Man Cybernetics, vol. SMC-8, 1978, pp. 86–92.
[20] Kanal, L. N.; and Chandrasekaran, B.: On Dimensionality and Sample Size in Statistical Pattern Recognition, Pattern Recognition, vol. 3, 1971, pp. 225–234.
[21] Kettig, R. L.; and Landgrebe, D. A.: Classification of Multispectral Image Data by Extraction and Classification of Homogeneous Objects, IEEE Trans. Geosci. Electron., vol. GE-14, 1976, pp. 19–26.

[22] Wiersma, D. J.; and Landgrebe, D.: The Use of Spatial Characteristics for the Improvement of Multispectral Classification of Remotely Sensed Data, Proceedings of Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Lafayette, Ind., 1976, pp. 2A-18–2A-22.

[23] Horwitz, H. M.; Nalepka, R. F.; Hyde, P. D.; and Morgenstern, J. P.: Estimating Proportions of Objects within a Single Resolution Element of a Multispectral Scanner. Seventh International Symposium on Remote Sensing of the Environment, Ann Arbor, Mich., May 1971.

[24] Chhikara, R. S.; and Odell, P. L.: Estimation of Proportions of Objects and Determination of Training Sample-Size in a Remote Sensing Application. Proceedings of Symposium on Machine Processing of Remotely Sensed Data, Purdue University, Lafayette, Ind., 1973, pp. 4B-16–4B-24.

[25] Fu, K. S., et al.: Information Processing of Remotely Sensed Agricultural Data, Proc. IEEE, vol. 57, 1969, pp. 639–653.

[26] Rodgers, E.; Siddalingaiah, H.; Chang, A. T. C.; and Wilheit, T.: A Statistical Technique for Determining Rainfall Over Land Employing Nimbus-6 ESMR Measurements. NASA TM 79631, Aug. 1978.

[27] Wilheit, T.: The Electrically Scanning Microwave Radiometer (ESMR) Experiment, *in* The Nimbus-6 User's Guide. NASA/Goddard Space Flight Center, Greenbelt, Md., Feb. 1975, pp. 87–108.

[28] Cochran, W. G.: The Chi-Square Test of Goodness of Fit, Ann. Math. Stat., vol. 23, 1952, pp. 315–345.

9. Image Data Compression

9.1 Introduction

The high data rates of multispectral remote sensing instruments create requirements for data transmission, storage, and processing that tend to exceed available capacities. These requirements are continually increasing with user demands for improved spatial and spectral resolution, Earth coverage, and data timeliness. For example, the future Landsat D system will transmit data from the Thematic Mapper instrument at a rate of 85 million bits per second and produce digital images in the form of 50 to 100 multi-images per day, each approximately 6,100 lines by 6,100 columns and containing seven spectral bands. One way to meet the requirements is through onboard data compression for transmission or through ground-based compression for archiving. Working with the image data in compressed form could reduce the cost of storage, dissemination, and processing [1].

A principal consideration in the decision to employ data compression is the effect on image fidelity. The criteria for image fidelity vary because many investigators use the data for several different purposes. According to figure 1.3, remotely sensed images are analyzed by machine and by subjective human evaluation. Therefore, fidelity measures should include diverse criteria, such as classification accuracy and properties of the human visual system. These properties are not clearly understood, and the performance criteria for machine analysis vary with the application.

In spite of considerable progress in data compression research, compression onboard spacecraft for transmission or on the ground for data dissemination and archiving has very seldom been used because of risks in reliability and data alteration [2].

Recent advances in system reliability and reduction of cost are making image data compression increasingly practical [3]. Image data compression can be accomplished either by exploiting statistical dependencies that exist between image samples or by discarding the data that are of no interest to the user. Thus, two types of compression, information-preserving and entropy-reducing compression, may be distinguished [4].

Information-preserving image compression is a transformation of an image for which the resulting image contains fewer bits. Hence, the original image can always be exactly reconstructed from the compressed image. The image statistics must be known to realize the transformation.

The stronger the correlation of the picture elements, the greater is the redundancy. Information-preserving compression removes this redundancy.

Entropy-reducing compression is an irreversible operation on the image and results in an acceptable reduction in fidelity. The operation depends also on the properties of the receiver. A compression acceptable in one application can be unacceptable in another. Examples are the image segmentation and classification techniques discussed in chapters 7 and 8, where the compressed data are represented by clusters or boundaries of regions in a scene. Entropy-reducing compression is in general not acceptable because of different user requirements. Only information-preserving compression techniques will be discussed in this chapter.

The redundancy in multi-images is due to the spatial correlation between adjacent picture elements and to the spectral or temporal correlation between the components of the multi-image. This redundancy can be modeled with image statistics and is therefore predictable. The output of the redundancy reduction step is called the derived data [5]. The nonuniform probability density of the remaining nonpredictable part of the image is a second source of redundancy that may be removed by coding [5, 6].

The purpose of image data compression is to remove the statistical predictability and then to encode the derived data for transmission or storage. Figure 9.1 shows the basic elements of an image data compression system. In the first step the redundancy due to the high correlation in the images is reduced. This redundancy reduction is a reversible process and thus preserves information. Compression techniques for this step include predictive compression, transform compression, and hybrid methods.

In the next step, the derived data are encoded with a fixed-length or variable-length code. Natural code, in which the data samples are represented in binary form, is a fixed-length code. The advantage of a fixed-length code is the constant word length. Its drawback is that redundancy due to the nonuniform data distribution still exists in the nonpredictable signal. This redundancy can be removed by using a variable-length code. A variable-length code maps short code words to the picture elements with higher probability of occurrence and long code words to rarely occurring data. After transmission or storage, the compressed data are decoded and reconstructed to images. To permit exact reconstruction, the nonredundant information content of the images must not be changed by compression.

The design of an image data compression system involves two basic steps. First, image properties and statistics, i.e., probability distribution functions for the gray values, entropy, and correlation functions, have to be calculated. The image statistics are used to determine information

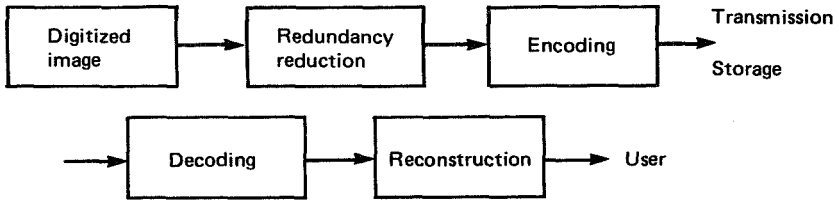


FIGURE 9.1—*Block diagram of image data compression system.*

content and redundancy and to model the predictive part of images. The next step is to define the compression technique and to establish performance criteria. The remainder of this chapter gives a summary of information-preserving image compression techniques.

9.2 Information Content, Image Redundancy, and Compression Ratio

The determination of the average information content of a class of images given by the entropy (see sec. 2.2.2) is an important requirement for data compression. The entropy of a picture element is defined as

$$H_e = -\sum_{k=1}^{n} p(k) \log_2 p(k) \tag{9.1}$$

where $p(k)$ is the probability of gray value k, and n is the number of members in the random field representing the images. The redundancy r is defined as

$$r = b - H_e \tag{9.2}$$

It can only be calculated if a good estimate of the entropy H_e is available. The quantity b is the number of quantization bits (see sec. 2.5.2) used to represent a picture element. The redundancy in digital remotely sensed images is due to the average information content of images being less than the number of bits used to represent them. Usually 6 to 10 bits per pixel are used on NASA remote sensing experiments.

Calculation of the entropy of a digital random field $\mathbf{g}$ is practically impossible. A field with M lines, N samples per line, and b bits per sample may generate $n = 2^{MNb}$ images. Calculation of the entropy with equation (9.1) requires knowledge of the probability of each image in the field. Therefore, the entropy may only be calculated on the basis of actual image data with the image histogram.

Given the histogram H_g of g, where $H_g(k)$ is the frequency of occurrence of gray level k and $0 \leq k \leq 2^b - 1$, the probability density of gray levels can be approximated by

$$p(k) = \frac{H_g(k)}{MN} \tag{9.3}$$

The entropy of the gray-level probability density is then

$$H_p = -\sum_{k=0}^{2^b-1} p(k) \log_2 p(k) \tag{9.4}$$

and the redundancy is

$$r_p = b - H_p \tag{9.5}$$

The compression ratio may be defined as

$$CR = \frac{b}{H_p} \tag{9.6}$$

Using the gray-level distribution results in an incorrect estimate of the entropy because of the correlation between gray levels. A better estimate for the entropy is obtained from the probability distribution of first gray-level differences

$$p(\Delta k) = \frac{H_g(\Delta k)}{MN} \qquad -(2^b-1) \leq \Delta k \leq 2^b-1 \tag{9.7}$$

where $H_g(\Delta k)$ is the frequency of gray-level difference Δk. The entropy of the probability distribution of first gray-level differences is

$$H_d(\Delta k) = -\sum_{\Delta k=-(2^b-1)}^{2^b-1} p(\Delta k) \log_2 p(\Delta k) \tag{9.8}$$

The representation of a digital multi-image consisting of P components, each given by an M by N matrix of picture elements with b bits per pixel, requires I bits with a conventional pulse-code modulation (PCM) code [7], where

$$I = bPMN \tag{9.9}$$

In a PCM code, each pixel value is represented by its b-bit binary number. The Multispectral Scanner (MSS) data of Landsats 1 and 2 are quantized to $b=7$ bits for bands 4, 5, and 6 and to $b=6$ bits for band 7. Given a frame size of 2,340 by 3,240 pixels, the total number of bits per multi-image is $I \approx 2 \times 10^8$. For a Landsat D Thematic Mapper multi-image with $b=8$, $P=7$, and $M=N \approx 6{,}100$, the total number of bits is $I \approx 2 \times 10^9$.

The entropies of the Landsat MSS image shown in figure 9.2 are listed in table 9.1. The entropy H_d is smaller than the number of quantization bits required by conventional PCM. Therefore, it should be possible to compress this image to an average of 4.2 bits per pixel with no loss of information and to achieve a compression ratio of 1.9.

9.3 Statistical Image Characteristics

For the analysis of compression techniques, it is desirable to have a model characterizing the image properties and involving only a few essential

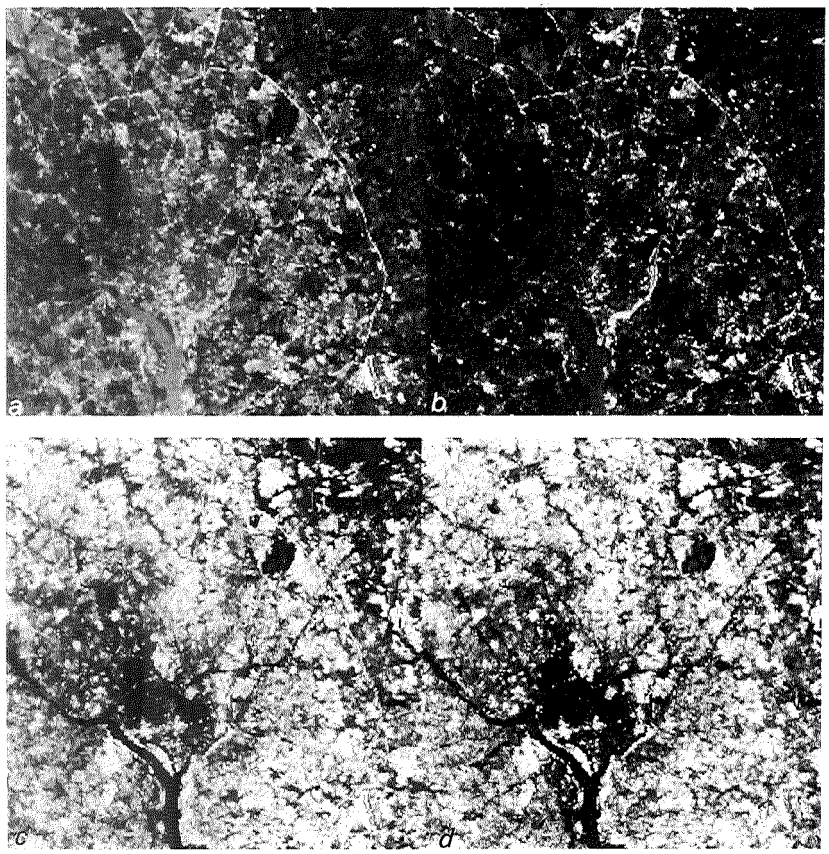


FIGURE 9.2—*Landsat multispectral image of Washington, D.C., area. (a) MSS 4. (b) MSS 5. (c) MSS 6. (d) MSS 7.*

TABLE 9.1—*Entropies of Landsat MSS Image in Figure 9.2*

MSS spectral band	Entropies[1]	
	H_p	H_d
4	4.15	3.94
5	4.4	4.38
6	4.86	4.63
7	4.28	3.99
Average	4.4	4.2

[1] H_p is the entropy of the gray-level probability density, given in equation (9.4); H_d is the entropy of the probability distribution of first gray-level differences, given in equation (9.8).

parameters. A useful model for multi-images is a random field $\mathbf{g}(x, y, \omega)$ (see sec. 2.2.2) where (x, y) are spatial variables and ω refers to the spectral or temporal variable. The spatial characteristics of a large class of remotely sensed images may be approximated by an autocorrelation function for a homogeneous random field $\mathbf{g}$ of the form [8]

$$R(\xi, \eta) = (R(0, 0) - \mu^2) e^{-\alpha|\xi|^2 - \beta|\eta|^2} + \mu^2 \tag{9.10}$$

This autocorrelation function depends only on the mean value μ, the variance $C(0, 0)$, where

$$C(0, 0) = R(0, 0) - \mu^2 \tag{9.11}$$

and the two parameters α and β, which specify the average number of statistically independent gray levels in a unit distance along the horizontal and vertical direction, respectively. In practice the autocorrelation function is computed as the spatial average given in equation (2.8). Figure 9.3 shows the horizontal and vertical autocorrelation functions of the image in figure 9.2, computed as averages of line and column correlation functions.

The correlation between spectral bands cannot be represented by the exponential model of equation (9.10). For example, Landsat MSS images often exhibit strong positive correlations between bands 4 and 5 and between bands 6 and 7 and small negative correlations between bands 5 and 7.

9.4 Compression Techniques

Although fine sampling and quantization are essential to preserve the subjective quality of a digital image, the information content could be conveyed with considerably fewer data bits. The approach taken in image compression is to convert the image data samples into a new set of uncorrelated variables that will contribute with a varying degree to the information content and subjective quality of the image. The less significant of these variables can then be discarded without affecting the information content and the subjective quality of the reconstructed image. This transformation to uncorrelated variables can be accomplished by prediction or by unitary transforms.

9.4.1 Transform Compression

Transform compression uses unitary transforms to remove the correlation of the data and to rank them according to the degree of significance to the information content of the image. The Karhunen–Loéve (K–L) transform (see sec. 2.6.1.4) results in a set of uncorrelated variables with monotonically decreasing variances. Because the information content of digital images is invariant under a unitary transform, and the variance of a variable is a measure of its information content, the compression

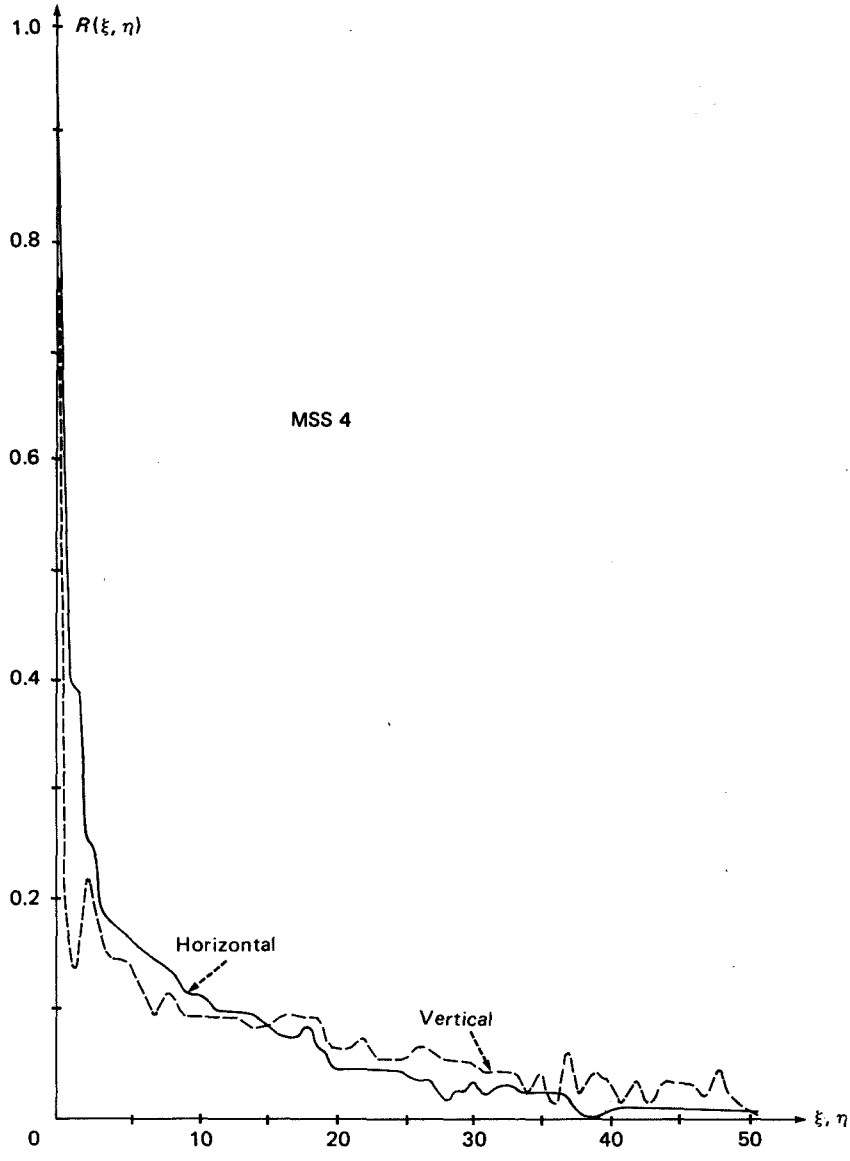


FIGURE 9.3a—*Average horizontal and vertical spatial autocorrelation functions of Landsat multispectral image in figure 9.2a.*

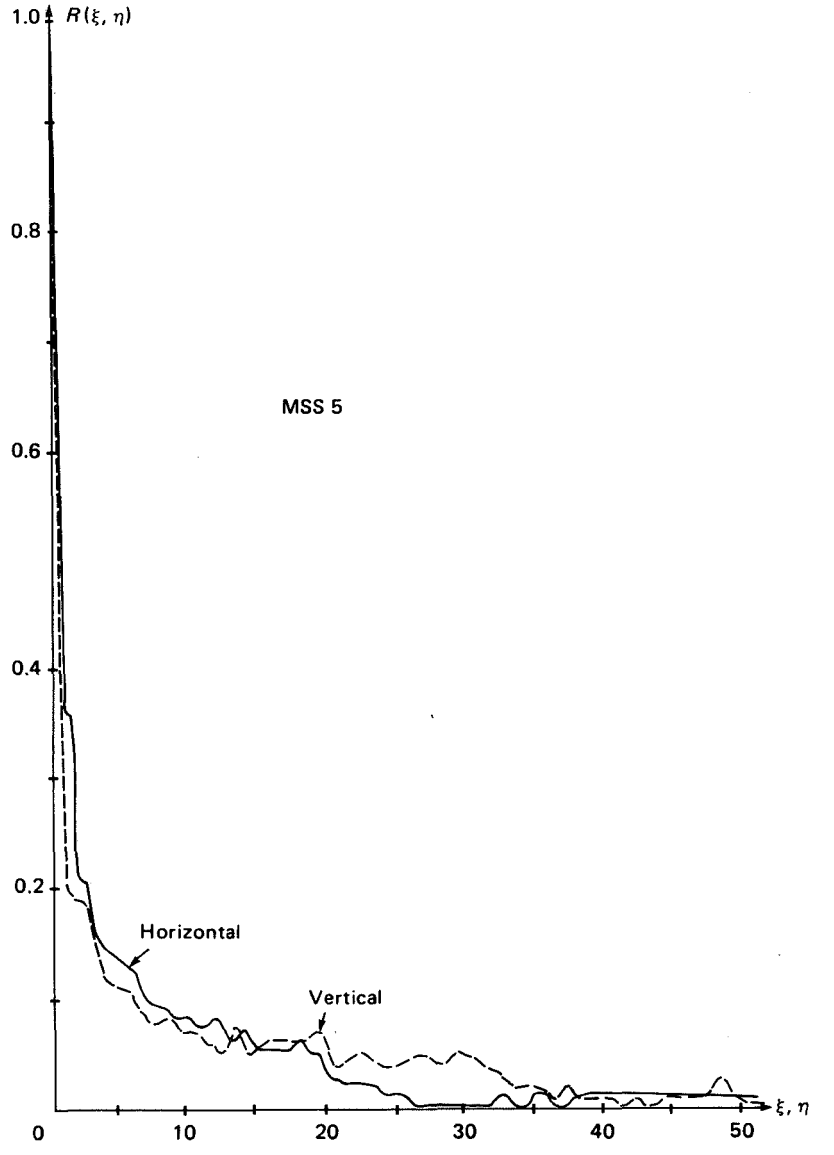


FIGURE 9.3b—*Average horizontal and vertical spatial autocorrelation functions of Landsat multispectral image in figure 9.2b.*

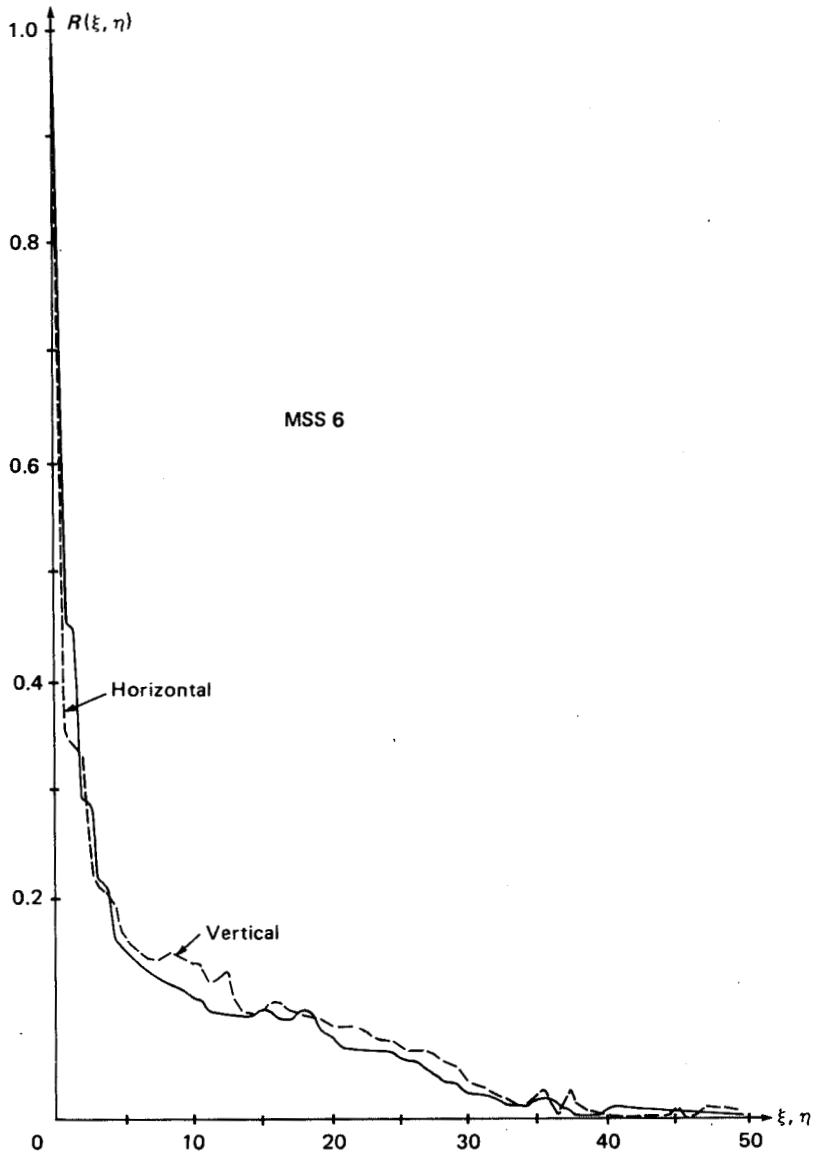


FIGURE 9.3c—*Average horizontal and vertical spatial autocorrelation functions of Landsat multispectral image in figure 9.2c.*

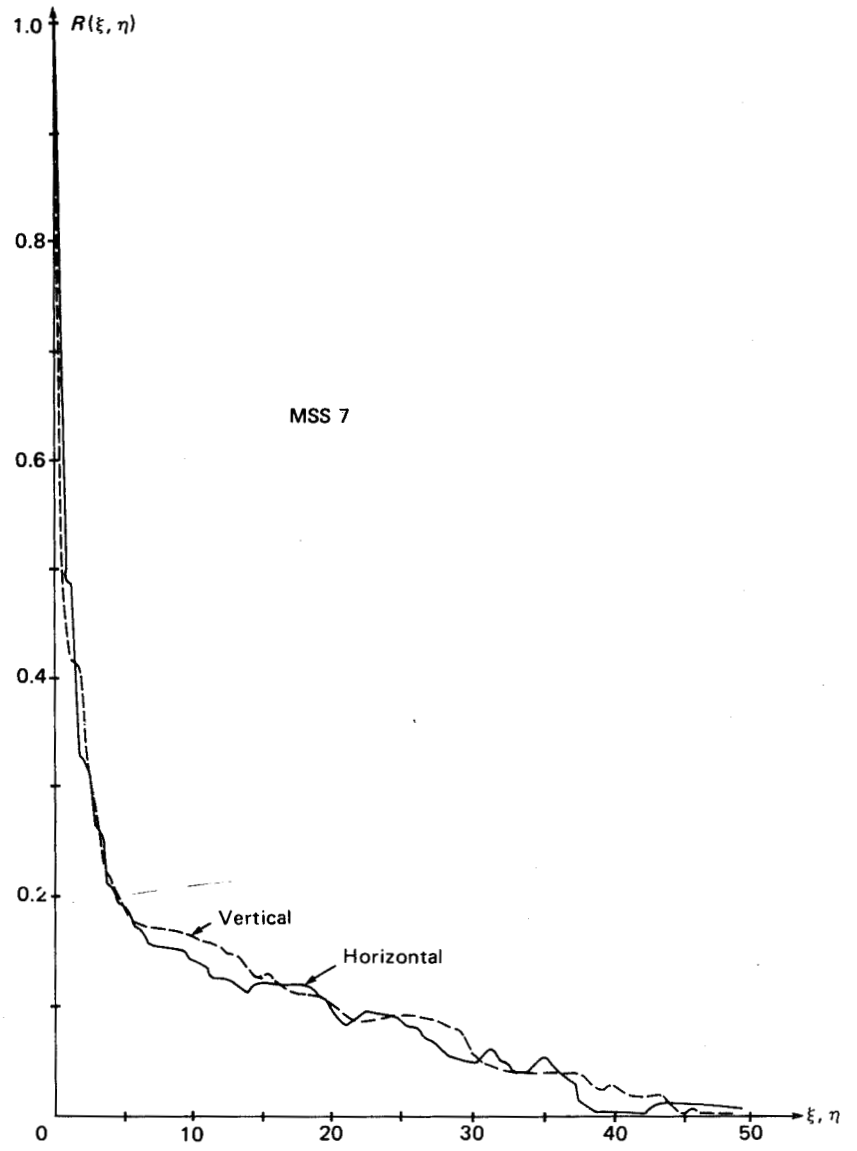


FIGURE 9.3*d*—*Average horizontal and vertical spatial autocorrelation functions of Landsat multispectral image in figure 9.2d.*

strategy is to discard variables with low variances [9–11]. The redistribution of variance in the principal components is important in an information–theoretic sense, because the K–L transform minimizes the entropy function defined over the data variance distribution [12].

The shortcomings of the K–L transform are that knowledge of the covariance matrix is required and that the computational requirements for the two-dimensional transform in the spatial domain are proportional to M^2N^2 for the forward and inverse transform. Furthermore, the eigenvalues and eigenvectors for the MN by MN covariance matrix have to be computed. The three-dimensional K–L transform for compression in the spatial and spectral dimensions is too complex to be considered. Therefore, only the one-dimensional K–L transform is applied in the spectral dimension where the correlation in general cannot be modeled by the exponential correlation function in equation (9.10).

The computation of the covariance matrix and its eigenvectors can be avoided if unitary transforms with a deterministic set of basis vectors are used. Such transforms are the Fourier, the cosine, and the Hadamard transforms. (See sec. 2.6.1.) Because of the existence of fast algorithms for these transforms, the computational requirements for the two-dimensional transformation are proportional to $MN \log_2 MN$ operations [9, 14]. The performance of these transforms is, however, inferior to the performance of the K–L transform, which is the only transform that generates uncorrelated coefficients. Only if the autocorrelation function is of the exponential form in equation (9.10) will the Fourier, cosine, and Hadamard transforms generate nearly uncorrelated coefficients. Because the spectral autocorrelation function of remotely sensed images is not exponential, these transforms are only employed in the spatial dimension.

Table 9.2 shows the spectral correlation matrix and its eigenvalues for the image in figure 9.2. The table also shows that 98.0 percent of the variance in the transformed data is contained in the first two components. Similar characteristics are shown in table 4.2 for aircraft scanner data.

9.4.2 Predictive Compression

Predictive compression uses the correlation between picture elements to derive an estimate $\hat{g}(i, j)$ for a given picture element $g(i, j)$ in terms of neighboring elements [15]. The difference $d(i, j) = \hat{g}(i, j) - g(i, j)$ between the estimate and the real value is quantized. Images are reconstructed by estimating values that are added to the differences. This technique is called differential pulse-code modulation (DPCM) [13].

Figure 9.4 shows the block diagram of a DPCM system. The transmitter is composed of a predictor and quantizer. The predictor uses n previous samples to predict the value of the present sample. The difference between this estimate and the actual value is quantized. The decoder in figure 9.4*b* reconstructs the image from the differential signal.

TABLE 9.2—*Spectral Correlation Matrix of Landsat MSS Image in Figure 9.2*

MSS spectral band	Spectral correlation matrix by MSS spectral band number				Eigenvalues	Percent variance
	4	5	6	7		
4	1.00	0.91	0.48	0.16	653.1	78.6
5	0.91	1.00	0.39	0.07	241.4	26.4
6	0.48	0.39	1.00	0.91	11.0	1.2
7	0.16	0.07	0.91	1.00	7.3	0.8

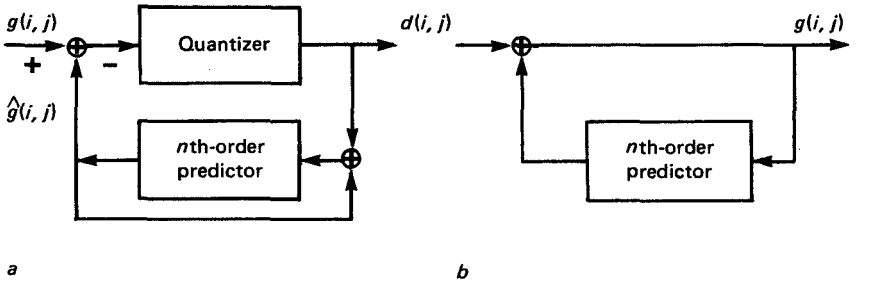


FIGURE 9.4—*Block diagram of DPCM system. (a) Generation of differential signal. (b) Reconstruction of image.*

Experimental results with various image data have indicated that a third-order linear predictor is sufficient to model digital images rather accurately [13]. With the assumption that the image is scanned row by row from top to bottom, the predicted value $\hat{g}(i, j)$ at line i and column j is then of the form

$$\hat{g}(i, j) = a_1 g(i, j-1) + a_2 g(i-1, j-1) + a_3 g(i-1, j) \qquad (9.12)$$

The predictor coefficients a_k, $k=1, 2, 3$ are determined such that the mean square error

$$\epsilon = E\{[g(i, j) - \hat{g}(i, j)]^2\} \qquad (9.13)$$

is minimized. For a homogeneous random field $\mathbf{g}$ with zero mean, the solution to this problem is obtained by solving the $n=3$ simultaneous equations

$$\left.\begin{aligned} a_1 R(0, 0) + a_2 R(0, 1) + a_3 R(1, 1) &= R(1, 0) \\ a_1 R(0, 1) + a_2 R(0, 0) + a_3 R(1, 0) &= R(1, 1) \\ a_1 R(1, 1) + a_2 R(1, 0) + a_3 R(0, 0) &= R(0, 1) \end{aligned}\right\} \qquad (9.14)$$

where $R(m, n)$ is the autocorrelation function of the random field $\mathbf{g}$.

For an exponential correlation function of the form given in equation (9.10), the variance of the differential data $d(i, j)$ is less than the variance of the picture elements $g(i, j)$, and the differential data are uncorrelated [13]. The predictor coefficients for the image in figure 9.2 are listed in table 9.3.

9.4.3 Hybrid Compression

Transform and predictive compression techniques generate uncorrelated or nearly uncorrelated data. Both unitary transforms and DPCM have advantages and limitations. Unitary transforms maintain subjective image quality better and are less sensitive to changes in image statistics than DPCM. On the other hand, DPCM achieves better compression at a lower cost [16].

Hybrid compression techniques combine the attractive features of both transform and predictive compression and avoid the limitations of each method [17]. Three categories of hybrid compression systems have been investigated [3]. First, two-dimensional transform compression of the individual components of a multi-image, followed by predictive compression data across the components, was used. Specifically, two-dimensional cosine and Hadamard transforms in combination with DPCM were evaluated. The second group involved one-dimensional transform compression in the spectral dimension with two-dimensional predictive compression in the transformed domain. Here, the combination K–L transform with two-dimensional DPCM was studied. The third group uses one-dimensional transforms in the spectral and horizontal spatial dimension and predictive compression in the vertical dimension. In this category, K–L/cosine/DPCM and K–L/Hadamard/DPCM combined transforms were selected for evaluation. The results showed that two-dimensional transforms with DPCM compression in the spectral dimension were inferior. This result occurred because spectral correlation is in general not exponential, and the number of spectral bands is usually small. A DPCM encoder does not reach a steady state, thus in this case causing inefficient performance. The two recommended hybrid com-

TABLE 9.3—*Predictor coefficients*

MSS spectral band	Horizontal			Vertical		
	a_1	a_2	a_3	a_1	a_2	a_3
4	6.337	0.089	0.071	0.072	0.197	0.129
5	0.315	0.078	0.065	0.154	0.146	0.083
6	0.396	0.078	0.071	0.251	0.181	0.082
7	0.432	0.072	0.081	0.328	0.172	0.059

pression techniques are K–L transform in the spectral dimension followed by one-dimensional cosine transform and DPCM, and K–L transform in spectral dimension followed by a two-dimensional DPCM.

9.5 Evaluation of Compression Techniques

To evaluate the performance of compression techniques, criteria measuring the distortion in reconstructed images must be defined. These criteria are necessarily application dependent. They include the mean square reconstruction error, signal-to-noise ratio (SNR), compression ratio, computational complexity, cost and error effects, subjective image quality, and influence on the accuracy of subsequent information extraction. For example, the effect of a particular compression technique on classification accuracy can be determined by comparing classifier performance on an original image and on the reconstruction of the compressed image.

9.5.1 Mean Square Error

The mean-square-error criterion is frequently used to compare the reconstructed image f with the original image g. Let $\epsilon(i,j)=g(i,j)-f(i,j)$ be the error at spatial coordinates (i,j). The mean square error e^2 is defined as

$$e^2=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}\epsilon(i,j)^2 \tag{9.15}$$

The error distribution is defined as

$$P(\epsilon)=\frac{H(\epsilon)}{MN} \tag{9.16}$$

where $H(\epsilon)$ is the frequency of occurrence of errors with magnitude ϵ. The peak error is defined as

$$\epsilon_{\max}=\max_{i,j}|\epsilon(i,j)| \tag{9.17}$$

and the average signal power s of the reconstructed image is

$$s=\frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}g(i,j)^2 \tag{9.18}$$

9.5.2 Signal-to-Noise Ratio (SNR)

The difference between the original and the reconstructed image may be considered as noise. Then the reconstructed image consists of the original image plus noise:

$$f=g+\epsilon \tag{9.19}$$

An SNR may be defined as the average power s of the reconstructed image divided by the mean square error:

$$\mathrm{SNR}_1 = 10 \log \frac{s}{e^2} \tag{9.20}$$

An alternate definition of an SNR is

$$\mathrm{SNR}_2 = 20 \log \frac{\Delta f_{\max}}{e^2} \tag{9.21}$$

where $\Delta f_{\max}$ is the maximum peak-to-peak signal value in the reconstructed image.

9.5.3 Subjective Image Quality

The properties of the human visual system are important considerations in the evaluation of compression techniques. The sensitivity of the human visual system depends logarithmically on light intensities that enter the eye. Thus, the higher the brightness f, the higher the contrast Δf between objects must be to detect any differences. This relationship is known as Weber–Fechner law. Furthermore, in the ability to detect fine spatial detail, the human visual system behaves like a band-pass filter. Thus, it is insensitive to the highest and lowest spatial frequencies in a scene. The importance of these nonlinear and spatial-frequency-dependent properties for image data compression was advocated by Stockham [18] and by Manos and Sakrison [19].

A simple psychophysical error (PE) criterion adapted to the nonlinear characteristic of the visual system can be defined [20] as a maximum:

$$\mathrm{PE}_m = \max_{i,j} \frac{|\epsilon(i,j)|}{f(i,j)+1} \tag{9.22}$$

or an average value:

$$\mathrm{PE}_a = \frac{1}{MN} \sum_{i=1}^{M} \sum_{j=1}^{N} \frac{|\epsilon(i,j)|}{f(i,j)+1} \tag{9.23}$$

These quantities are calculated for each component of a multi-image and the average over all components is used to evaluate compression techniques for multi-images.

REFERENCES

[1] Lynch, T. J.: Data Compression Requirements for the Landsat Follow-On Mission. NASA Goddard Space Flight Center, Report X-930-76-55, Feb. 1976.

[2] Miller, W. H.; and Lynch, T. J.: On-Board Image Compression for the RAE Lunar Mission, IEEE Trans. Aerosp. and Electron. Syst., vol. AES-12, 1976, pp. 327–335.

[3] Habibi, A.: Study of On-Board Compression of Earth Resources Data. TRW Report CR137752, Sept. 1975.

[4] Blasbalg, H.; and VanBlerkom, R.: Message Compression, IRE Trans. Space Electron. Telemetry, vol. 8, 1962, pp. 228–238.
[5] Chen, P. H.; and Wintz, P. A.: Data Compression for Satellite Images. Tech. Report TR-EE 77-9, School of Electrical Engineering, Purdue University, Lafayette, Ind., 1976.
[6] Capon, J.: A Probabilistic Model for Run Length Coding of Pictures, IRE Trans. on Inf. Theory, vol. IT-5, 1959, pp. 157–163.
[7] Huang, T. S.: PCM Picture Transmission, IEEE Spectrum, vol. 2, Dec. 1965, pp. 57–63.
[8] Franks, L. E.: A Model for the Random Video Process, Bell Syst. Techn. J., vol. 45, Apr. 1966, pp. 609–630.
[9] Wintz, P. A.: Transform Picture Coding, Proc. IEEE, vol. 60, 1972, pp. 809–820.
[10] Habibi, A.; and Wintz, P. A.: Image Coding by Linear Transformation and Block Quantization, IEEE Trans. Commun. Technol., vol. COM-19, 1971, pp. 50–62.
[11] Pratt, W. K.; Kane, J.; and Andrews, H. C.: Hadamard Transform Image Coding, Proc. IEEE, vol. 57, 1969, pp. 58–68.
[12] Watanabe, S.: Karhunen-Loéve Expansion and Factor Analysis, Theoretical Remarks and Applications. Transactions of the Fourth Prague Conference on Information Theory, Prague, Czechoslovakia, 1965.
[13] Habibi, A.: Comparison of nth Order DPCM Encoder with Linear Transformations and Block Quantization Techniques, IEEE Trans. Commun. Technol., vol. COM-19, no. 6, 1971, pp. 948–956.
[14] Anderson, G. B.; and Huang, T. S.: Piecewise Fourier Transformation for Picture Bandwidth Compression, IEEE Trans. Commun. Technol., vol. COM-19, 1971, pp. 133–140.
[15] Elias, P.: Predictive Coding, IRE Trans. Inf. Theory, vol. IT-1, 1965, pp. 16–23, 30–33.
[16] Habibi, A.; and Robinson, G. S.: A Survey of Digital Picture Coding, IEEE Comput., vol. 7, 1974, pp. 22–35.
[17] Habibi, A.: Hybrid Coding of Pictorial Data, IEEE Trans. Commun., vol. COM-22, 1974, pp. 614–623.
[18] Stockham, T. G.: Intra-frame Encoding for Monochrome Images by Means of a Psychophysical Model Based on Nonlinear Filtering of Signals. Proceedings of 1969 Symposium on Picture Bandwidth Reduction, Gordon and Breach Sci. Pub., New York, 1972.
[19] Manos, F.; and Sakrison, D. L.: The Effects of a Visual Fidelity Criterion on the Encoding of Images, IEEE Trans. Inf. Theory, vol. IT-20, 1974, pp. 525–536.
[20] Bruderle, E., et al.: Study on the Compression of Image Data Onboard an Applications or Scientific Spacecraft. Report for ESRO/ESTEC Contract 2120/73 HP, Mar. 1976.

Symbols

a	altitude sensor, aperture radius
$\mathbf{a}$	weight vector, separating vector
a_d	gain
A	amplitude of periodic pattern
$\mathbf{A}$	transformation matrix
b	word length of natural binary code, number of quantization bits
b_d	offset
B	path radiance
$B(j, k)$	brightness image
$\mathbf{C}$	covariance matrix
C_{fg}	covariance function of f and g
C_k	covariance matrix for class S_k
δ	Dirac delta function
$\Delta\lambda_j$	spatial distance for a given spectral band
$\nabla^2 g$	Laplacian of image g
d_P	degree of match between two images
d_A	absolute difference between images
d	Mahalanobis distance, interclass distance
D	number of detectors in sensors, edge density
D_A	average divergence
$D(S_j, S_k)$	divergence between classes of S_j and S_k
D_T	transformed divergence
$\mathbf{D}$	spacecraft orientation matrix
e	error
e_{mn}	mean square error
ϵ	expected mean square error
E	expectation operator
$e(j, k)$	edge image
η	spatial integration variable
$\mathbf{f}$	multi-image, random field
$\hat{f}$	restored image matrix
$\hat{f}$	restored image, estimate of original image of f
$\hat{\mathbf{f}}$	vector estimate of f
j_f	training pattern
$f^*(x, y)$	apparent object radiant energy
$\mathbf{f}(x_0, y_0)$	multidimensional picture element, pixel

$\mathcal{F}$	Fourier transform operator
$F_{\mu v}$	transform coefficients
F^*	complex conjugate of F
$F(0)$	discrete cosine transform (DCT)
$F(0, 0)$	two-dimensional DCT
$F(m, n)$	Hadamard transform
$F(u, v)$	Fourier transform of $f(x, y)$
g	recorded image
g_d	reconstructed display image
g_f	filtered image
g_R	radiometrically degraded image
g_s	digital or sampled image
g_t	threshold image
g_e	enhanced image
g_0	image radiant energy
$g_k(z)$	linear discriminant function
g	convolution of two functions
$\mathbf{g}^c$	principal component image
$g_k{}^D$	difference image
$g_k{}^R$	ratioed image
$\|G(u, v)\|$	magnitude of Fourier transform $G(u, v)$
G_f	Fourier transform of filtered image
G_s	Fourier transform or frequency spectrum of g_s
G_d	frequency spectrum of display image
$G(u, v)$	Fourier transform of $g(x, y)$
$G(\mathbf{w})$	Fourier transform (vector form)
$g(\mathbf{z})$	inverse Fourier transform (vector form)
h_0	optical system point spread function
h_d	impulse response of display interpolation filter
h_e	edge spread function
h_l	line spread function
h_t	truncated filter impulse response
$h_s(x, y)$	sampling impulse
$h(x, y)$	point spread function (PSF) of linear space-invariant imaging system, filter impulse response
$H(j, k)$	hue image
$H_g(t)$	histogram of image g
$H_I(u, v)$	inverse filter
H_s	Fourier transform of h_s
H_d	transfer function of display interpolation filter
H_e	entropy
$H_e(\mathbf{z})$	histogram of enhanced image

$H(u, v)$	Fourier transform of h, optical transfer function (OTF), filter transfer function
i	imaginary unit
${}^{i}\mathbf{f}$	training pattern
$i(x, y, \lambda, t)$	irradiance
I	identity transformation
ρ	correlation coefficient
J	Jacobian
$J(a)$	criterion function
$J_0(r)$	zero-order Bessel function
$J_1(w)$	first order Bessel function
K	number of clusters, maximum gray value
K_a	number of quantization levels
Λ	diagonal matrix
λ	longitude, wavelength
$\lambda\ (S_i, S_k)$	loss function
L	radiance
$L(\mathbf{z}, S_k)$	average loss
$\mathcal{L}$	linear system operator
μ	mean value
$\mathbf{m}$	mean vector
m_d	mean gray value
$\mathbf{m}_k$	mean vector for class S_k
$M(u, v)$	modulation transfer function (MTF)
$M(\lambda)$	spectral radiant emittance of a blackbody
$\mathbf{M}$	spacecraft rotation matrix
M_k	number of training vectors for class S_k
n	noise
n_r	random noise
n_s	structured noise
n	normal from satellite to Earth surface, yaw axis
N	number of features
$N(u, v)$	Fourier transform of noise n
Ω	set of all events
φ	latitude
$\phi(y)$	phase
$\phi_{\mu v}$	complete set of orthonormal functions
$\Phi(u, v)$	phase of Fourier transform, phase transfer function (PTF)
p	polarization
$p(x, y), q(x, y)$	coordinate transformation functions
$\mathbf{p}$	spacecraft position vector
$p(z)$	probability density function for feature vectors

$p(\mathbf{z})$	mixture probability density
$p(\mathbf{z}/S_k)$	probability density distribution of $\mathbf{z}$ for class
p_e	estimate of probability of correct classification
p_i	probability of event ω_i
p_e	error rate of a classifier
p_f	joint probability density
PE	psychological error criterion
P	dimensionality of multi-image
$P(\epsilon)$	error distribution
P_k	probability distribution
$P(S_k)$	a priori class probability
$P(S_i\|\mathbf{z})$	a posteriori probability
$\mathbf{Q}$	kernel matrix
$r(x, y, \lambda, t, p)$	reflectance
r	redundancy
R	rectangular region
R	image domain
R_i	region in image domain
$\mathbf{R}_k$	correlation matrix for class S_k
$\mathbf{R}$	correlation matrix
R_{ff}	autocorrelation function of f
R_{fg}	crosscorrelation function of f and g
$R(m, n)$	crosscorrelation metric
R_s	statistical correlation measure
$R(u, v)$	homomorphic restoration filter
σ	standard deviation
σ^2	variance
σ_p^2	variances of the principal components
σ_d	standard deviation
s	average signal power, training set
$\mathbf{s}$	scanner pointing direction
$s(j, k)$	saturation image
s_k	training set for class S_k
$s(x, y)$	sampling function
$\mathbf{S}$	satellite spin axis, roll axis
S_R	reject class
S	set of all patterns
S_k	spectral densities
S_{ff} and S_{gg}	subset of S, pattern class
$S(m, n)$	similarity measure
$S(u, v)$	Fourier transform of sampling function
τ	spectral transmittance
θ	angle of scan-mirror deflection
t	time

$\mathbf{t}_n$	orthonormal vectors
T	image recording time
T	threshold
T	texture image
$\mathbf{T}_e$	transformation from input image to geodetic coordinates
$\mathbf{T}_p$	map projection transformation
$\mathbf{T}_s$	sealing transformation
$\mathbf{T}$	transformation matrix
$\mathbf{T}_g$	geometric transformation
T_G	geometric distortion transformation
T_i	image degradation transformation
$\mathbf{T}_n$	reduced transformed matrix
T_R	radiometric degradation transformation
u, v	spatial frequencies
U, V	frequency limits of a band-limited function
$\mathbf{v}$	spacecraft velocity vector
V_x	sensor velocity
V_n	volume of N-dimensional hypersphere
$w(x, y), w(\tau)$	window-function
$\mathbf{W}$	weight vector
$W(u, v)$	Wiener filter
$W(u, v), W(\omega)$	Fourier transform of w
$W(\omega)$	Fourier transform
ω_c	cutoff frequency
ω	radial spatial frequency
ω_i	an event
(x', y')	image coordination system
(X', Y')	object coordinate system
T_c	gray-scale transformation
ξ	spatial integration variable
$\mathbf{z}$	feature vector

Page intentionally left blank

GLOSSARY OF IMAGE PROCESSING TERMS

Acutance: Measure of the sharpness of edges in an image.

Aliasing: Image misrepresentation and/or loss of information due to undersampling. Overlap of frequency spectra in sampled images.

Aspect ratio: Ratio between scales in horizontal and vertical direction.

Band-pass filter: Spatial filter that suppresses spatial frequencies outside a specified frequency range.

Bayes decision rule: Decision rule that treats the patterns independently and assigns a pattern or feature vector $\mathbf{c}$ to the class S_k whose conditional probability $P(S_k \mid \mathbf{c})$, given pattern $\mathbf{c}$, is highest.

Change detection: Process by which two images are compared pixel by pixel, and an output is generated whenever corresponding pixels have sufficiently different gray values.

Cluster: Homogeneous group of patterns that are very similar to one another as determined by the distance between patterns or by their density.

Clustering: Process that assigns patterns to a cluster on the basis of the training patterns.

Clustering algorithm: Function that is applied to the unlabelled training patterns to yield a sequence of clusters.

Color composite: Color image produced by assigning colors to three selected compounds of a multi-image.

Conditional probability density: Probability density of a pattern or feature vector $\mathbf{c}$ given class S_k, denoted by $p(\mathbf{c} \mid S_k)$ and defined as the relative number of times the vector $\mathbf{c}$ is derived from an image area whose true class identification is S_k.

Control points: Recognizable geographic features or landmarks in an image that relate the image with the object.

Decision boundary: Boundary between classes S_i and S_k in pattern space. (It may be thought of as a subset H of pattern space S such that $H = \{\mathbf{c} \in S \mid g_i(\mathbf{c}) = g_k(\mathbf{c})\}$, where g_i and g_k are the discriminant functions for classes S_i and S_k.)

Decision rule: Rule in which one and only one class is assigned to each observed pattern or feature on the basis of a training set.

Density slicing or thresholding: Operation to assign a range of gray values to one value.

Digitization: Partitioning of an image into discrete resolution cells, then assignment of a representative gray value to each cell.

Dirac delta function: An ideal pulse of infinite height and zero width whose integrated area is equal to unity.

Discriminant function: Scalar function $g_k(\mathbf{c})$ whose domain is pattern or feature space, whose range is the class numbers, and which assigns class number S_k to a pattern $\mathbf{c}$.

Edge: Abrupt change in brightness in an image.

Edge spread function: Response of a linear space-invariant imaging system to an edge input.

False color: See color composite.

Feature: Vector whose components are functions of the initial measurement patterns. Gray value, texture measure, or coefficient of orthogonal transform.

Feature selection: Process by which the features used in classification are determined from the measurement patterns.

Feature space: The set of all possible feature vectors.

Filter or spatial filter: An image transformation that assigns a gray value at location (x, y) in the transformed image on the basis of gray values in a neighborhood of (x, y) in the original image.

Frequency spectrum: Function representing the image components for each spatial frequency. Formally the magnitude of the Fourier transform of an image.

Geometric distortion: Distortion due to movements of either sensor, platform, or object.

Gray scale: Range of gray values from black to white.

Gray shade or gray value: Number that is assigned to a position (x, y) on an image and that is proportional to the integrated image value (reflectance, radiance, brightness, color coordinate, density) of a small area, called a resolution cell, or a pixel, centered on the position (x, y).

High-pass filter: Spatial filter that suppresses low spatial frequencies in an image, and enhances fine detail.

Histogram: Function representing the frequency of occurrence of gray values in an image.

Homomorphic filtering: Nonlinear filtering that transforms the problem by a logarithmic operation to linear filtering.

Hyperplane decision boundary: Decision boundary arising from the use of linear discriminant functions.

Image: Spatial representation of an object, a scene, or a map, which may be abstractly represented by a continuous function of two variables defined on some bounded region of a plane.

Image classification: Process, often preceded by feature selection, in which a decision rule is applied that assigns class numbers to unknown patterns or features on the basis of the training set.

Image compression: Operation that reduces the amount of image data and the time needed to transmit an image while preserving all or most of the information in the image.

Image enhancement: Improvement of the detectability of objects or patterns in an image by contrast enhancement, edge enhancement, color enhancement, or multi-image enhancement.

Image processing: All operations that can be applied to image data, including preprocessing, image restoration, image enhancement, image registration, image segmentation, sampling, quantization, image classification, and image compression.

Image registration: Alignment process by which two images of the same scene are positioned coincident with respect to each other so that corresponding elements of the same ground area appear in the same position on the registered images.

Image restoration: Operation that restores a degraded image to its original condition.

Image segmentation: Operation to determine which regions or areas in an image constitute objects or patterns of interest.

Image transformation: Operation that takes an image as input and produces an image as output. (The transform operator's domain is the spatial domain and its range is the transform domain. For the Fourier and Hadamard transformations, for example, the transform domain has an entirely different character from the spatial domain. For the Karhunen–Loéve transform or filtering transformations the image in the transformed domain may appear similar to the image in the spatial domain.)

Line spread function: Response of a linear space-invariant imaging system to a line input.

Low-pass filter: Spatial filter that suppresses high spatial frequency components in an image, and suppresses fine detail or noise.

Map: Representation of physical and/or cultural features of a region or a surface such as that of Earth, indicating by a combination of symbols and colors those regions having designated category identifications. Representation displaying classification category assignments.

Maximum likelihood decision rule: Decision rule that treats the patterns independently and assigns a pattern or feature vector $\mathbf{c}$ to that class S_k that most probably gave rise to pattern or feature vector $\mathbf{c}$; that is, such that the conditional probability density of $\mathbf{c}$ given S_k, $p(\mathbf{c} \mid S_k)$, is highest.

Modulation transfer function: Function that measures the spatial frequency modulation response of a linear imaging system and indicates for each spatial frequency the ratio of the contrast modulation of the output image to the contrast modulation of the input image.

Mosaic: Combination of registered images to cover an area larger than an image frame.

Multi-image: Set of images, each taken of the same scene at different times, or at different electromagnetic wavelengths, or with different sensors, or with different polarizations.

Multispectral image: Multi-image whose components are taken at the same time in different spectral wavelengths.

Multitemporal image: Multi-image whose components are taken in one spectral wavelength at different times.

Nonparametric decision rule: Decision rule that makes no assumptions about the functional form of the conditional probability distribution of the patterns given the classes.

Notch filter: Inverse of a band-pass filter; suppresses all frequencies within a given band of spatial frequencies.

Nyquist frequency: One-half the sampling rate ($\frac{1}{2}\ \triangle$), two samples per cycle of the Nyquist frequency being the highest observable frequency.

Pass points: Recognizable features in an image that relate a series of images.

Pattern class or category: Set of patterns of the same type.

Pattern or pattern vector: The ordered n-tuple or vector of measurements obtained from a resolution cell. (Each component of the pattern measures a particular property.)

Picture element or pixel: The gray value of a resolution cell in an image, or the gray values of a resolution cell in a multi-image.

Pixel: See picture element.

Point spread function: Response of a linear space-invariant imaging system to a point light source.

Polarization: Restriction of vibration of a transverse wave to a single direction—the direction of the electric vector in a light wave.

Preprocessing: Operation applied before image analysis or image classification is performed, which can remove noise from, bring into registration, and enhance images.

Pseudocolor: Assigning colors to specific ranges of gray values in an image.

Quantization: Process by which a gray value or a range of gray values in an image is assigned a new value from a given finite set of gray values.

Radiometric degradation: The effects of atmosphere and imaging systems that result in a blurred image. Degradation resulting from nonlinear amplitude response, vignetting, shading, transmission noise, atmospheric interference, variable surface illumination, etc.

Radiometric resolution: The sensitivity of the sensor to the differences in signal strength, defining the number of discernible signal levels.

Reflectance: Ratio of the energy per unit time per unit area reflected by an object to the energy per unit time per unit area incident on the

object. A function of the incident angle of the energy, viewing angle of the sensor, spectral wavelength and bandwidth, and the nature of the object.

Resolution cell: The smallest elementary areal constituent of gray values considered in an image, referred to by its spatial coordinates.

Resolving power of an imaging system: An imaging system's ability to image closely spaced objects, usually measured in line pairs per millimeter, i.e., the greatest number of lines and spaces per millimeter that can just be recognized.

Roll: Rotation about the velocity vector causing panoramic distortion.

Sampling: Process of measuring the brightness or intensity of a continuous image of discrete points, producing an array of numbers.

Signature: The characteristic patterns or features derived from units of a particular class or category.

Skew: Distortion due to the rotation of the spacecraft about the local zenith vector (yaw).

Spatial resolution: A description of how well a system or image can reproduce an isolated object or separate closely spaced objects or lines in terms of the smallest dimension of the object that can just be observed or discriminated.

Spectral bands: An interval in the electromagnetic spectrum defined by two wavelengths, frequencies, or wave numbers.

Spread function: A description of the spatial distribution of gray values produced by a linear imaging system when the input to the system is some well-defined object.

Template matching: Operation that determines how well two images match each other by crosscorrelating the two images or by evaluating the sum of the squared gray value differences of corresponding pixels.

Temporal resolution: The time interval between measurements.

Training set: Sequence of pattern subsets, $s=(s_1, \ldots, s_k)$ such that s_k is derived from class k, which is used to estimate the class conditional probability distributions from which the decision rule may be constructed.

Transmittance: Ratio of the energy per unit time per unit area transmitted through an object to the energy per unit time per unit area incident on the object.

Vidicon: Vacuum tube with a photosensitive surface.

Vignetting: Gradual reduction in density of parts of a photographic image caused by preventing some of the rays from entering the lens.

Page intentionally left blank

Index

A Posteriori Probability, 189, 266, 267
A Priori Probability, 1–2, 255, 266, 269
Aberration, 36, 43
Acuity, 70
Adaptive Threshold Selection, 226
Additive Noise, 89
Affine Transformation, 19, 108
Aliased Spatial Frequency, 57
Aliasing, 47, 68, 69, 113
AOIPS—*see* Atmospheric and Oceanographic Image Processing System
Aperture, 21, 43
Apodization, 57
Archiving, 293
Artificial Contour, 149
Artificial Edge, 200, 227
Aspect Ratio, 89
Atmosphere, 4, 5, 77–79
Atmospheric
  Correction, 78–79, 158
  Effect, 2, 4–5, 34, 43, 77–78, 128
  Motion, 235
  Transmittance, 4–5, 34, 36, 38, 212
Atmospheric and Oceanographic Information Processing System, 27, 235
Attenuation, 127
  Atmospheric, 5, 34
  Spatial Frequency, 113
Attitude, 36, 41
  Determination, 107
  Errors, 36, 41–42
  Precision, 107
  Time Series, 107
Autocorrelation, 298
  Of a Function, 12, 13, 21, 298
  Of a Random Field, 13, 15
Average
  Divergence, 262
  Ensemble, 15
  Loss, 274
  Transformed Divergence, 284
Averaging, 89, 298
Background, 70
Band Interleaved by Line, 45
Band Interleaved by Pixel, 46
Band-Limited Function, 47, 57, 68, 110
Band-Pass Filter, 70, 80, 82, 153
Band Sequential, 45
Bandwidth, 38
Battacharyya Distance, 254
Bayesian Classifier, 266, 268–269, 273, 282, 284
Bayesian Rule, 267–268
Bessel Function, 16, 23, 43
Bilinear Interpolation, 111
Binary Correlation, 227
Binary Image, 164
Binary Mask, 164
Binomial Distribution, 282
Bipolarized Measurement, 285
Bit (Unit of Information), 13, 49
Bivariate Polynomial, 108, 196, 211
Blur
  Atmospheric, 34, 44
  Motion, 34
Blurring, 27, 42, 44, 113, 120, 127, 130, 134, 137
Border, 70, 224
Boundary, 270–276
Boundary Detection, 227
Brightness, 42, 70–72, 148–149
  Component, 78, 148
  Difference, 72
  Temperature, 149
  Variation, 48, 128, 148
Calibration, 79, 153
Camera Frame, 34
Camera Error, 77
  Aberration, 36, 43
  Shading, 78, 193
  Vignetting, 42
Cartographic Projection, 89, 103, 200
Cauchy-Schwartz Inequality, 190
Census Tract, 237, 243
Change Detection, 187, 199, 211, 243
Channel, 77–78, 127
Chi-Square
  Distribution, 269
  Test, 268, 290
Chromatic Variation, 149
Circulant Matrix, 64
Circular (Cyclic) Convolution, 21
Circular Symmetric Filters, 27
Class
  Boundary, 251
  Characteristics, 251
  Discrimination, 255
  Separability, 249–284
  Separation—*see* Classification

Classifier
Bayesian, 266, 268, 273, 282, 284
Evaluation, 281–284
Hybrid, 293, 305–306
Implementation, 249–253
Linear, 253, 274
Maximum Likelihood, 284
Optimal, 268
Parallel, 277
Parallelepiped, 271–273
Quadratic—*see* Bayesian Classifier
Sequential, 277
Classification
Accuracy, 78, 261
Algorithm, 251, 254
Error, 78
Geometric, 273–278
Nonparametric, 266
Parametric, 266
Time, 270, 277
Unsupervised, 251, 278
Supervised, 251, 263, 285
Statistical, 266–273
Cloud
Displacement, 103, 235
Height, 199, 235
Cluster, 279, 280
Clustering, 237, 278–281
Algorithm, 280
Coefficients of Expansion, 254
Coherent Noise, 37, 77, 79
Color
Assignment, 148–149
Difference, 72
Display, 149
Distribution, 148–149
Enhancement, 72, 127
Information, 149–158
Parameter, 127, 148
Perception, 127
Primary, 141, 159
Space, 48, 72, 149
Variation, 149
Color Composite, 149, 153, 158, 159
Compass Course, 206
Component
Image, 148
Illumination, 34, 44
Of a Multi-Image, 9, 10
Principal, 63, 82, 164, 170–171, 176
Reflectance, 34
Composite Image
Edge Image, 27, 196
Image Color, 149, 153, 158, 159
Compression, 63, 69
Entropy Reducing, 293–294
Evaluation, 305–306
Hybrid, 294, 305–306
Information Preserving, 293–305
Predictive, 294, 303–305
Transform, 294, 298–303
Conditional
Average Loss, 266–267
Probability Density, 266
Contour, 70
False, 49, 52
Contrast
Attenuation, 127
Characteristics, 70, 153, 192
Enhancement, 49, 127–130, 149, 153, 192
Convolution, 11–12, 15, 24–26, 56, 57, 63–65, 120, 137
Circular (Cyclic), 196
Theorem, 15, 47, 64
Coordinate Transformation, 103–109, 203
Correlation, 194–196
Coefficient, 195
Function, 13, 122, 196, 298
Matrix, 60–61, 63, 256
Measure, 191, 256, 298
Peak, 194
Surface, 194
Cosine Transform, 55, 59–60
Covariance
Function, 14, 252, 253
Matrix, 170, 194, 195, 252, 256–259
Total, 256
Convolution Integral, 12
Criterion Function, 274, 276–277, 279–280
Crosscorrelation
Normalized, 191
Matching, 12, 190, 191, 196, 235
Of a Function, 12, 19–20, 21, 65–67
Of a Random Field, 14
Cumulative Distribution Function, 153
Cutoff Frequency, 27, 113
Data Compression, 293–307
Datum Surface, 201–203
Deblurring, 120, 127, 130, 134, 137
Decision
Boundary, 286
Maximum Likelihood, 266–268
Region, 260, 270
Rule, 251, 267, 268
Surface, 274
Theory, 273
Degradation, 39–46, 68
Atmospheric, 2, 4–5, 34, 43, 77–78, 128
Geometric, 34, 36, 40–42, 77, 89, 199–222
Motion, 43
Point, 40
Radiometric, 38, 77
Temporal, 44

Delta Function, 11–12, 69
Density
  Edge, 227
  Normal, 252, 258
  Spectral, 79, 122, 129
  Slicing, 130, 149
Derived Data, 294
Description of Images, 223, 234
Determinant, 252
Deterministic Restoration, 118
Detector, 39, 77, 87
DFT—*see* Discrete Fourier Transform
Diagonalization of Matrix, 63
Difference Image, 153, 159, 164
Differencing, 70, 159–164, 191, 243
Differentiation, 130
Differential Pulse Code Modulation, 303–305
Diffraction Effects, 43
Digital Data Format
  BIP—*see* Band Interleaved by Pixel
  BIL—*see* Band Interleaved by Line
  BSQ—*see* Band Sequential
Digital Image, 25–49, 110
Digital Image Processing, 2, 6, 49
Digitization, 2, 44–46
Dimensionality Reduction, 254–263
Dirac Delta Function, 11, 12, 23, 46
Directional Difference, 225
Discrete Fourier Transform, 55–59
Discriminant Function, 267, 270, 274
Discrete Transform, 52–69
Discrete Approximation to Convolution, 63–65
Displacement, 108, 191
Display, 67–69
  Device, 67, 128, 148–149
  Nonlinearity, 130, 148
  Systems, 69, 148–149
Distance
  Battacharyya, 254
  Euclidian, 279
  Interclass, 255, 258
  Mahalanobis, 252
  Mean-Square, 255, 258, 304, 306
  Measure, 279
Distortion
  Geometric, 40–42, 192
  Map, 200–201
  Measure, 187–306
  Radiometric, 199
Distribution
  Binomial, 282
  Chi-Square, 269
  Normal, 268
  Parameter, 282
  Of a Function, Normal, 252, 268
Divergence
  Average, 261–262
  Pairwise, 260, 261
  Transformed, 284
DPCM—*see* Differential Pulse Code Modulation
Dynamic Range, 128
Earth Coordinates—*see* Geodetic Coordinates
Easting, 203
Edge, 21, 69
  Density, 227
  Detection, 27, 225–232
  Detector, 226
  Effect, 69, 200, 212
  Enhancement, 72, 127, 130, 133–134, 137, 153, 158, 225–227
  Following, 227
  Image, 27, 196
  Spread Function, 119
Eigenvalues, 62, 120, 170–171
Eigenvectors, 62, 170–171, 252, 255, 256
Electro-Optical System, 37
Electrically Scanning Microwave Radiometer, 285
Emission, 37
Encoder, 305
Encoding, 294
Energy, 10
Enhancement, 12, 70, 167, 170–171
  Color, 72, 127
  Contrast, 49, 70–72, 127–130, 149, 153, 192
  Directional, 134
  Edge, 72, 127, 130, 133–134, 137, 153, 158, 225–227
  False Color, 149, 159, 164
  Filter, 34, 35, 63, 123, 137, 141, 153
  Image, 23, 63
  Multi-Image, 158–164, 187
  Pseudocolor, 72, 149
Ensemble Average, 15
Entropy, 13, 295–296
Entropy Reducing Compression, 293
Ergodic, 15
Ergodic Random Field, 15
Error Rate, 281
ESMR—*see* Electrically Scanning Microwave Radiometer
Euclidean Distance, 279
Event, 13
Expansion, 89
Expansion—Orthogonal, 60
Expansion—Orthonormal, 61, 62, 255
Expected Value, 252
Eye, 148
False Contours, 49, 52
False Color, 149–158
  Enhancement, 159, 164
  Image, 113

Fast Fourier Transform, 56, 196
Feature
  Evaluation, 259, 284
  Extraction, 254
  Selection, 63, 251, 253–254, 255
  Space, 271
  Subset, 259–262
  Vector, 251–252, 273
FFT—*see* Fast Fourier Transform
Fidelity Criteria, 149, 294
Field of View, 2, 199
Film Characteristics, 128–130
Filter, 25–26
  Bandpass, 25
  Design, 17, 27
  Exponential, 25
  Gaussian, 25, 27
  High-Pass, 25, 27, 34, 130
  Homomorphic, 141
  Inverse, 25, 118, 120–121
  Linear, 24, 25
  Low-Pass, 25, 27, 113
  Notch, 25
  Wedge, 27
  Wiener, 118
Filtering, 24–34
  Homomorphic, 34, 35, 123, 141, 153
  Edge, 137
  Linear, 63
Foreshortening, 40
Forest Insect Damage, 243, 282
Fourier
  Compression, 303
Fourier Transform, 11, 15, 17–21, 24–27, 55, 254, 255
  Discrete, 47, 55–56, 82, 120
  Inverse, 17, 18, 56, 210
  Noise Removal, 80–82
  Properties, 15, 17–21, 55–58, 80
FOV—*see* Field of View
Frame, Image, 34, 80, 212, 296
Fredholm Integral Equation, 119
Frequency
  Spatial, 17, 25
  Spectrum, 88
  Domain, 25, 88
Gaussian Density—*see* Density, Normal
Gaussian Filter, 25, 27
GCP—*see* Ground Control Point
Geodetic Coordinates, 103–108, 209
Geometric
  Classification, 273–278
  Correction, 200
  Distortion, 34, 36, 40–42, 77, 89, 199–222
  Errors, 40–42, 192
  Transformation, 89, 103–119
Geosynchronous Orbit, 235
GEOS—*see* Geostationary Operational Environmental Satellite
Geostationary Operational Environmental Satellite, 104, 107, 235
Gibbs Phenomenon, 25
Gradient, 134
  Descent, 277
  Digital, 225
Gray Value, 10
Gray Level, 45, 128, 224
Gray Scale, 11
  Difference, 149
  Normalization, 153, 266
  Transformation, 128, 129
Ground Control Point, 210, 235–237
Hadamard
  Compression, 303
  Matrix, 303
  Transform, 55, 60, 303, 305
Hankel Transform, 16, 21–23
Hanning Window, 26
Haze, 34, 78
HCMM—*see* Heat Capacity Mapping Mission
Heat Capacity Mapping Mission, 149
High Pass Filter, 27, 34
Histogram
  Bimodal, 130
  Equalization, 130
  Flattening, 130, 153, 159
  Gaussian, 130, 153
  Of Gray Levels, 79, 129
  Modification, 130
  Multidimensional, 130
  Normalization, 153, 226
HO–Kashyap Algorithm, 277
Homogeneity, 232 (*also, see* Classification)
Homogeneous Random Field, 13, 15, 23
Homomorphic Filter, 34, 153
Homomorphic Filtering, 34, 35, 123, 140
Hue, 149
  Scale, 72, 148, 149
Human Visual System, 49, 69–73, 127, 149, 293, 307
Hybrid Classifier Compression, 293, 305–306
Hyperellipsoid, 257, 271
Hyperrectangle, 271
Hypersurfaces, 273
Hyperplane, 273, 274
IBIS—*see* Image Based Information System
Ideal Filter, 121–122
IFOV—*see* Instantaneous Field of View
Illumination, 77–78
Illumination Component, 34, 44

Image
Analysis, 223–246
Characteristics, 232–234
Classification, 234, 249–290
Data Compression, 63, 69, 293–306
Description, 223, 234
Digital, 45
Display System, 119
Enhancement, 23, 127–171, 223
Entropy, 13, 295, 296
Fidelity, 293
Formulation, 34, 36, 37, 44, 45
Mosaic, 78, 199–200, 212, 219
Multi, 9, 10
Monochromic, 128, 149
Multispectral, 10, 49, 149
Multitemporal, 10, 282, 283, 284
Overlay, 89, 103, 199–200, 211–212
Plane, 37
Properties, 63, 232, 249, 253, 294, 296
Quality, 36, 307
Recording, 34, 36, 37, 38, 44, 45
Registration, 89, 187–196, 235
Restoration, 77–123, 223
Statistics, 294
Segmentation, 223–234, 243
Transform—*see* Transform
Reconstruction, 294
Image-Based Information System, 237
Image Processing—*see* Digital Image Processing
Image Radiant Energy, 9, 37
Impulse Response, 24, 68
Information Content, 12, 13, 295–296, 298
Extraction, 293
Information Preserving Compression, 293
Insect Infestation, 243, 282
Instantaneous Field of View, 2, 199
Intensity, 36, 109, 110, 149, 212, 224, 307
Intensity Errors, 192–194
Interactive Image Analysis System, 6–7
Interactive Image Processing System, 6
Interclass Distance, 255, 258
Interference Pattern, 79–80
Interpolation Grid, 107, 109–112
Interpolation
Bicubic, 111–113
Bilinear, 111–113, 210
Nearest Neighbor, 110–111, 113
Inverse Filter, 25, 120–121
Inverse Filtering, 25, 118, 120–121
Inverse Transform, 17, 18, 56, 210
Jacobian, 19
Jacobian Determinant, 108
Karhunen—Loéve Transform, 16, 23, 55, 164, 170–171, 176, 254–255, 298
Kernel Matrix, 255–256
Land Use, 237, 243
Landmark—*see* Ground Control Point
Landsat, 38, 49
Landsat MSS, 49, 87, 88, 107, 111, 113
Laplacian, 134, 137
Latitude, 106, 207
Leakage, 57
Least Squares
Estimate, 120
Filtering, 120
Legendre Polynomial, 108
Level Slicing, 149
Lexicographical Ordering, 52
Line Spread Function, 119
Linear
Classifier, 253, 274
Discrimination Function, 267, 270, 274
Inequality, 276
Operation, 17–19, 23, 38
System, 2, 23–24, 38
Transform, 15, 23, 108, 128–129, 254, 259
Linear Approximation to Imaging Equation, 111
Linearized Imaging Model, 77
Linearly Separable, 275
Linear Shift
System, 19
Invariant Operation, 20
Longitude, 106
Loss Function, 269
Loss Function—Symmetric, 267
Low-Pass Filter, 25, 27, 113
Loxodrome, 206–207
Luminance, 70, 78, 153
Mahalanobis Distance, 252
Map Projection, 209
Map Overlay, 211
Map Matching, 78, 199–200, 212, 219
Markov Process, 63
Mask—Binary, 164
Matching, 190, 192
Match Point, 191
Matrix
(Block) Circulant, 64
Covariance, 61–63, 170, 194, 195, 252, 256–259
Hadamard, 303
Orthogonal, 55, 106, 255
Singular, 60
Symmetric, 60, 252
Transpose, 57, 62
Unitary, 55
Maximum Likelihood
Algorithm, 268–269
Classifier, 270
Decision Rule, 267

Mean
  Of a Random Field, 15
  Of a Random Variable, 249
  Of a Class, 252, 258, 259, 273
Mean Square Error, 16, 119, 121, 122, 255, 304, 306
Mean Vector, 252, 268, 280
Mercator Projection, 205–207
Meridian, 205
Misregistration, 187, 196
Modulation
  Pulse Code, 49, 296
  Transfer Function, 24, 44, 72, 113, 121
Moire Pattern, 69
Mosaic, 78, 199–200, 212, 219
Motion
  Blur, 34, 235
  Degradation, 43
MSDS—*see* Multispectral Scanner Data System
MSS—*see* Multispectral Scanner
MTF—*see* Modulation Transfer Function
Multi-Image, 9, 10, 13
Multi-Image Enhancement, 158–164, 187
Multispectral
  Classification, 281–284
  Image, 10, 49, 149
  Information, 149–158
  Scanner, 171
Multivariate Normal Density, 252, 258 (*also, see* Density, Gaussian)
Multiplicative Noise, 85
Multispectral Image, 10, 49, 149
Multispectral Scanner Data System, 171
Multitemporal Image, 10, 282, 283, 284
Multispectral Sensor, 38, 49
Munsell Color System, 72
Nadir Point, 78, 209
Nearest Neighbor Interpolation, 110–111
Nimbus, 104, 286
Noise
  Additive, 40
  Coherent, 37, 77, 79
  Film Gain, 37
  Multiplicative, 85
  Random, 37, 77, 114
  Removal, 79–89, 119, 170–171
  Transmission, 79
  Uncorrelated, 37, 89
Nonlinear Contrast Enhancement, 130
Nonlinear Sensor Characteristics, 87
Nonparametric Classification, 266
Normal Distribution, 252, 268
Normal Mercator Projection, 205–207
Normal Probability Density Distribution, 273
Normal Projection, 203–204
Northing, 203
Notch Filter, 80, 88
Nyquist Frequency, 47, 59, 69, 118
Nyquist—Shannon Theorem, 110
Object Plane, 37
Object Radiant Energy, 2, 36, 37, 38, 77
Object Distribution, 37, 114, 118
Oblique Mercator Projection, 207
Oblique Projection, 207
Operation
  Linear, 17–19, 23, 38
  Shift Invariant, 19–20
Optical Mechanical Scanner, 34
Optical System, 38, 40–42
Optical Transfer Function, 24, 43, 44
Optimal Classifier—*see* Bayesian Classifier
Optimal Filter, 121–122
Optimal Quantization, 128
Orthogonal
  Expansion, 60
  Function, 55
  Matrix, 55, 106, 255
  Transformation, 63, 254–259
  Vectors, 255
Orthonormal
  Basis, 16
  Function, 15, 16
  Expansion, 61, 62, 255
OTF—*see* Optical Transfer Function
Overlay, 89, 103, 199–200, 211–212
Pairwise Divergence, 260, 261
Panoramic Distortion, 40–41
Parallelepiped Classifier, 271–272
Parametric Classification, 266
Parseval's Theorem, 116
Partitioning, 278
Path Radiance, 36, 37, 78–79
Pattern
  Clusters, 279
  Recognition, 187, 279
  Space, 249, 257–258, 279
  Vector, 249, 257–258
PCM—*see* Pulse Code Modulation
Perception, 148
Performance Index, 279
Performance of a Classifier, 281–284
Periodic Noise, 79–80
  Point, 40, 43
  Projection, 42, 203–205
Phase Transfer Function, 27, 34
Phase, 34
Photomosaic, 78, 199–200, 212, 219
Photochemical Image Detection and Recording, 37

Picture Element—*see* Pixel
Piecewise Linear Transformation, 128–129
Pitch, 41, 104, 105
Pixel, 38
Planar Projection, 201
Point
  Degradation, 40, 43
  Source, 11, 24
  Spread Function, 24, 37, 40–44, 77, 114, 119, 130, 134
Polarization, 2, 9–10, 286
Polarstereographic Projection, 199
Polar Coordinates, 148
Polyhedric Projection, 201
Polycylindrical Projection, 201
Polyconic Projection, 201
Polysuperficial, 201
Polynomial
  Orthogonal, 55, 60
  Bivariate, 108
Positive Restoration, 119, 120
Power Spectrum, 21, 88, 232–233
Predictor, 303–304
Predictive Compression, 294, 303–305
Primary Color, 141, 159
Principal Component, 63, 82, 164, 170–171, 176
Principal Component—Image, 63, 82, 164, 170–171, 176
Probability
  A Posteriori, 189, 266, 267
  A Priori, 255, 266, 269
  Of Correct Classification, 259
Probability Density, 13, 252, 260, 263, 269, 294
  Multivariate Normal, 252, 258
Probability Distribution, 13, 87, 294
Projection
  Axis 207, 263
  Center, 201
  Conformal, 203, 205, 207
  Conical, 200–201
  Cylindrical, 200–201, 205
  Lambert, 207–208
  Mercator, 205–207
  Normal, 201, 204
  Oblique, 201
  Perspective, 203–205
  Planar, 201
  Plane, 200, 209
  Polyconic, 201
  Polycylindrical, 201
  Polyhedric, 201
  Polystereographic, 204, 205
  Polysuperficial Projection, 201
  Secant, 201
  Surface, 200–201
  Tangent, 201
  Transverse, 201
  UTM, 208–209
Property of an Object or Image, 232, 249, 253, 294, 296
PSF—*see* Point Spread Function
Pseudocolor
  Enhancement, 149, 159
  Image, 148–149
  Transformation, 148–149
Psychophysical Error, 307
PTF—*see* Phase Transfer Function
Pulse Code Modulation, 49, 296
Quadratic Classifier—*see* Bayesian Classifier
Quadratic Form, 269
Quality, 298
Quantization, 34, 45, 49–52, 128
  Bits, 41, 295
  Error, 47
  Linear, 49, 128
  Noise, 2
  Nonlinear, 49, 128
  Nonuniform, 49, 128
  Optimal, 128
  Uniform, 49, 128
Quantizer, 303
Radiant Energy, 9, 37
  Image, 37
  Object, 34
Radiometric
  Correction, 200
  Degradation, 34, 42–44, 114
  Difference, 200
  Error, 38, 77
  Restoration, 12, 24–34, 69, 114–123
  Transformation, 103–109
Random
  Field, 11, 12–15, 295, 298
  Noise, 12
  Variable, 12–15, 249, 266
Random Field
  Ergodic, 15
  Homogeneous, 13, 15, 16, 23
  Independent, 13
  Uncorrelated, 13
Raster Image, 45, 199
Ratio Image, 78, 85, 153, 158
Ratioing, 158–159
RBV—*see* Return Beam Vidicon
Recorded Image, 37, 38, 40, 41
Reconstruction, 67–69
Reconstructed Image, 69, 294
Reconstruction Filter, 118, 121
Recognition Accuracy—*see* Classification Accuracy
Redundancy, 13, 293
Redundancy Reduction, 294–305

Reference Image, 192–194
Reference Grid, 200
Reflectance, 77
Reflectance Component, 34
Region, 9, 224, 227
Registration
  Algorithm, 189–194
  Correlation, 190–192, 194–196
  Error, 191–192
  Filter, 194
  Image, 89, 187–196, 235
  Procedure, 187
  Statistical Correlation, 187
Reject
  Class, 269
  Region, 270
Remotely Sensed Images, 2, 9–11, 79
Resampling, 107, 109–114
Resolution
  Element, 11
  Radiometric, 11
  Spatial, 11, 39, 286
  Spectral, 11
  Temporal, 11
Restoration
  Constrained, 119
  Deterministic, 118
  Inverse Filter, 120–121, 122
  Least-Squares Filter, 118
  Mean-Square-Error, 121–122
  Positive, 119, 122
  Stochastic, 118
  Wiener Filter, 118, 121
Return Beam Vidicon, 4, 38, 77
Ringing, 26, 27, 113
Roll, 41, 104, 105
Rotation, 20, 40–41, 89, 103, 108–109, 192
Sampling, 44–49, 68
  Error, 47
  Function, 46–49, 57
  Grid, 46, 49
  Interval, 41, 45, 46, 49, 57
  Theorem, 47
Saturation, 72, 129, 149, 159
Scale Errors, 192, 208
Scaling, 20, 108, 109, 279
Scanning Imaging System, 34–36, 40–41, 127
Scan Angle, 78
Scan Geometry, 40–41, 80, 103–107, 196
Scanner Orientation, 40–41
Scattering, 34, 37, 78, 79
Seam, 200
Search Area, 188–194
Segmentation of an Image, 223–234, 243
Sensor, 2, 9, 199
Sensor Characteristics, 36, 40–41, 78, 87, 192, 199
Sensor Response, 85, 88, 128
Separability of Classes, 249–284
Separable Linearly, 276, 277
Separating Vector, 276
Severe Storm, 4, 235
Shading, 34, 42, 78, 128
Shape, 70, 127
Sharpening, 134, 141
Sharpness, 119
Shifting Property, 19, 66–67
Signal-to-Noise Ratio, 49, 79, 120, 122, 193, 306
Signature Extension, 78
Significance of a Cluster, 279, 280
Similarity Measure, 190, 191, 193, 278–279
Skew Errors, 192
Skewing, 41, 89, 108, 109
Small Interactive Image Processing System, 6–7
SMIPS—*see* Small Interactive Image Processing System
Smoothing, 89, 219
SMS—*see* Synchronous Meteorological Satellite
SNR—*see* Signal-to-Noise Ratio
Space Invariant System, 24, 37, 122, 130
Space Invariant–Point Spread Function, 24, 37, 40–44, 77, 114, 119, 130, 134
Spacecraft
  Altitude, 103–105
  Attitude, 39, 106–107
  Position, 103–105
  Velocity, 38, 103–104
Spatial
  Average, 89, 298
  Coordinates, 103–109, 203
  Correlation, 195
  Distortion, 42, 44
  Domain, 12, 13, 25, 26
  Frequency, 21, 24, 25, 47, 69, 80, 113
  Frequency—Aliased, 57
  Frequency Attenuation, 113
  Frequency—Spectrum, 9–10
  Registration, 187–196
Spectral
  Band, 4, 13
  Component, 78
  Difference, 44, 63, 158
  Domain, 9
  Emittance, 9
  Irradiance, 9
  Radiance, 9
  Reflectance, 9
Spike Noise, 79, 89
Spin-Stabilized Spacecraft, 104

Spread Function
  Edge, 119
  Line, 119
  Point, 24, 37, 40–44, 77, 114, 119, 130, 134
Standard
  Deviation, 191
  Parallel, 207
Statistical
  Analysis, 12
  Characterization of Images, 12–16, 60
  Correlation, 194–196
  Classification, 266–273
  Predictability, 294, 303–305
Stereographic Projection, 199
Stochastic Restoration, 118
Striping, 79, 85–88, 269
Structural Description of an Image, 223, 234
Subimage, 188, 249
Subjective Interpretation, 140
Sun Angle, 77
Supervised Classification, 251, 263, 285
Superposition, 194
Symmetric Matrix, 252
Synchronous Meteorological Satellite, 39, 87, 104
Systematic Intensity Errors, 193
Table Lookup, 69, 270
Television
  Display, 2
  Camera, 2, 34
Template, 188, 194–195
Template Matching, 188–192, 194, 195
Temporal Change, 2, 11, 44, 63, 164, 200, 243
Test Pattern, 71
Test Set, 282, 284
Texture, 73, 127, 224
  Analysis, 232–234
  Edge Detection, 225, 232–233
  Feature, 233
  Measure, 233
  Property, 225, 232–234
Thematic Mapper, 293, 296
Thermal Infrared Image, 149
Thresholding, 164, 224–225, 243, 269
Threshold Selection, 226
  Adaptive, 226
Tie-Point, 108
Topographic Relief, 158
Training
  Area, 251–263
  Pattern, 251–252
  Vector, 252
Training Set, 250, 255, 266, 268, 282
  Selection, 251–263
Tracking, 235–237
Transfer Function, 24, 141
Transform
  Compression, 294, 298–303
  Cosine, 55, 59–60
  Discrete, 52–55
  Discrete—Fourier, 55–59
  Discrete—Cosine, 59–60
  Discrete—Karhunen–Loéve, 60
  Fourier, 11, 15, 17–21, 303
  Hadamard, 60, 303
  Hankel, 16, 21–23
  To Principal Components, 164, 170–171
  Unitary, 15–17, 55–58, 298, 303
Transformation, 1
  False Color, 149–158
  Geometric, 103–109
  Gray Scale, 130, 153, 226
  Linear, 15–17, 279
  Orthogonal, 63, 254–259
  Pseudocolor, 148–149
  Rotational, 279
  Scaling, 108, 109, 279
  To Principal Components, 164, 170–171, 176
Transformed Divergence, 284
Transmission, 34, 127
Transmittance—Atmospheric, 4–5, 34, 36, 38, 212
Transverse Mercator Projection, 208–209
Tristimulus Value, 72
Uncorrelated Random Fields, 14, 296–305
Unitary Transform, 15–17, 55–58, 298
Universal Transverse Mercator Projection, 208–209
Unsupervised Classification, 251, 278
Urban Atlas File, 237
UTM Projection—*see* Universal Transverse Mercator Projection
Variance, 63, 164, 171, 252, 259, 298
VICAR—*see* Video Image Communication and Retrieval System
Video Image Communication and Retrieval System, 6
Video Infrared Spin Scan Radiometer, 87, 107
Vidicon Camera, 1, 38
Viewing Geometry, 77
Vignetting, 34, 42, 128
VISSR—*see* Video Infrared Spin Scan Radiometer

Visual Perception, 69–70
Visual System, 69–73
Web Grammar, 234
Weber–Fechner Law, 70, 307
Weight Space, 275
  Vector, 275, 276
Wiener Filter, 118, 121–122
Wind
  Vector, 235
  Field, 4, 234–237
  Velocity, 235
Window Hanning, 26
Wraparound, 65
Yaw, 41, 104, 105

☆ U.S. GOVERNMENT PRINTING OFFICE: 1980 O–296-053

National Aeronautics and Space Administration

Washington, D.C. 20546

Official Business
Penalty for Private Use, $300

SPECIAL FOURTH CLASS MAIL BOOK

Postage and Fees Paid
National Aeronautics and Space Administration
NASA-451

U.S. MAIL

7 1 1U,E, 040681 S00903DS
DEPT OF THE AIR FORCE
AF WEAPONS LABORATORY
ATTN: TECHNICAL LIBRARY (SUL)
KIRTLAND AFB NM 87117

NASA

POSTMASTER: If Undeliverable (Section 158 Postal Manual) Do Not Return